



**Universidade de
Aveiro**
2015

Departamento de Electrónica, Telecomunicações
e Informática

Cátia Matos Carvalho

**Plataforma Web para apoio às Direções de Curso:
análise do desempenho nas unidades curriculares**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Dr.-Ing. Joaquim João Estrela Ribeiro Silvestre Madeira, Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro e coorientação científica do Doutor José Manuel Matos Moreira, Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

o júri

presidente

Professor Doutor Paulo Miguel de Jesus Dias
Professor Auxiliar, Universidade de Aveiro

Professora Doutora Maria Beatriz Duarte Pereira do Carmo
Professora Auxiliar, Universidade de Lisboa - Faculdade de Ciências

Professor Dr.-Ing. Joaquim João Estrela Ribeiro Silvestre Madeira
Professor Auxiliar, Universidade de Aveiro

agradecimentos

Agradeço aos meus pais e aos meus irmãos pelo apoio e motivação constantes.

Agradeço também a todos os meus amigos e colegas pela ajuda que me foram dando e pelos bons momentos que me proporcionaram ao longo destes últimos anos.

Gostaria ainda de agradecer ao Professores Joaquim Madeira e José Moreira pela orientação neste trabalho de dissertação.

palavras-chave

Visualização de Informação, Plataforma Web, Desempenho Académico

resumo

É importante para os diretores de curso terem acesso a dados que permitam determinar o desempenho dos alunos nas unidades curriculares por que são responsáveis mas devido à grande quantidade de informação, torna-se difícil analisar esses dados.

A utilização de métodos da Visualização de Informação para representar esses dados facilita a sua compreensão e análise.

Esta dissertação tem como objetivo gerar visualizações que representem o número de alunos inscritos, avaliados e aprovados nas diversas unidades curriculares dos cursos do Departamento de Electrónica, Telecomunicações e Informática, assim como as taxas de abandono e aprovação nessas mesmas unidades curriculares, para que os diretores de curso possam determinar facilmente o desempenho dos alunos nas unidades curriculares que lhes dizem respeito.

É apresentada uma introdução aos sistemas de suporte à decisão e à Visualização de Informação descrevendo o que são e que ferramentas e tecnologias podem ser utilizados na sua implementação e descrita a implementação da plataforma desenvolvida.

keywords

Information Visualization, Web Plataform, Academic Performance

abstract

It is important for the course directors to have access to data that allow measuring the performance of students in curricular units which they are responsible for but, due to the large amount of information, it is difficult to analyze such data.

The use of Information Visualization techniques to represent such data makes its understanding and analysis easier.

This dissertation aims to generate visualizations representing the number of students enrolled, evaluated and approved in the various curricular units of the courses of the Department of Electronics, Telecommunications and Informatics, as well as drop-out and approval rates for those curricular units so the directors of the course can easily determine the performance of students.

An introduction to decision support systems and information visualization is done, describing what they are and what tools and technologies can be used in its implementation, and the implementation of the developed platform is described.

Índice

1. Introdução	1
1.1 Motivação e Contexto	1
1.2 Objetivos.....	1
1.3 Estrutura da dissertação	2
2. Estado da Arte	3
2.1 Sistemas de Suporte à Decisão.....	3
2.1.1 Metodologias de Desenvolvimento	3
2.1.2 Modelos de Dados Multidimensionais	7
2.1.3 OLAP – Processamento Analítico de Dados	7
2.1.4 Ferramentas e Tecnologias	8
2.2 Visualização de Informação	9
2.2.1 Processo de visualização de informação	10
2.2.2 Ferramentas e Tecnologias	11
2.3 Sistemas de apoio à decisão na educação e visualização de dados educacionais	14
2.4 Sumário.....	15
3. Desenvolvimento	17
3.1 Requisitos funcionais.....	17
3.2 Atores	17
3.3 Modelo de casos de uso	17
3.4 Descrição dos casos de uso	18

3.5 Estrutura do sistema.....	19
3.6 Organização dos Dados	21
3.7 Resultados	25
4. Avaliação	33
4.1 Avaliação Heurística.....	33
4.1.1 Heurísticas de Nielsen.....	33
4.1.2 Heurísticas de Zuk et al.	35
4.1.3 Avaliação	35
4.2 Testes de Usabilidade	37
5. Conclusões e trabalho futuro	41
5.1 Resumo	41
5.2 Trabalho futuro.....	42
6. Bibliografia	43
Apêndice A Testes de Usabilidade: Documentos	45
Apêndice A.1 Guia de tarefas do Utilizador	45
Para o gráfico de barras disponível a partir da página inicial.	47
Para o gráfico de linhas disponível a partir da página inicial.	48
Apêndice A.2 Ficha do Observador	49

Índice de Figuras

Figura 1 – Esquema de armazenamento de dados num Data Warehouse.....	4
Figura 2 – Metodologia de Kimball [9]	5
Figura 3- Metodologia de Inmon. Adaptado de [13]	6
Figura 4 – Processo de visualização de informação [21].....	11
Figura 5 – Gráfico de barras construído utilizando a biblioteca Google Charts [22]	12
Figura 6 – Gráfico de barras construído utilizando a biblioteca Highcharts JS [23].....	12
Figura 7 – Gráfico construído utilizando a biblioteca Charts JS [24].....	13
Figura 8 – Gráfico construído utilizando a biblioteca d3.js [25].....	13
Figura 9 – Representação multicamada[26]	14
Figura 10 – Representação multimatricial [26].....	15
Figura 11 – Modelo de casos de uso	18
Figura 12 – Modelo de casos de uso do utilizador.....	19
Figura 13 – Modelo de casos de uso do sistema.....	19
Figura 14 – Esquema da estrutura MVC.....	20
Figura 15 – Diagrama de Componentes.....	21
Figura 16 – Modelo da Base de Dados Multidimensional.....	24
Figura 17 – Excerto do ficheiro Excel com os dados agregados.....	25
Figura 18 – Página de entrada da plataforma.....	26
Figura 19 – Página do gráfico de barras sem filtros ativos.....	27
Figura 20 – Gráfico de barras	27
Figura 21 – Etiqueta apresentada no gráfico de barras.....	28
Figura 22 – Mensagem informativa na página do gráfico de barras	28
Figura 23 – Gráfico de barras na horizontal.....	29
Figura 24 – Página do gráfico de linhas.....	29
Figura 25 – Gráfico de linhas.....	30
Figura 26 – Gráfico de linhas para apenas um curso	30
Figura 27 – Etiqueta apresentada no gráfico de linhas.....	31
Figura 28 – Tempo de execução das tarefas por parte de cada utilizador	38
Figura 29 – Dificuldade indicada por cada utilizador na execução das tarefas	39

Acrónimos

BI – Business Intelligence

ETL – Extract-Transform-Load

HOLAP – Hybrid On Line Analytical Processing

HTML – Hyper Text Markup Language

JSON – JavaScript Object Notation

MOLAP – Multidimensional On Line Analytical Processing

MVC – Model-View-Controller

OLAP - On Line Analytical Processing

PDO – PHP Data Objects

PHP – PHP: Hypertext Preprocessor

ROLAP – Relational On Line Analytical Processing

SQL – Structured Query Language

1. Introdução

1.1 Motivação e Contexto

No ambiente académico, as direções de curso sentem necessidade de ter acesso a dados relativos às unidades curriculares do curso por que são responsáveis com o objetivo de detetar eventuais situações anómalas e perceber padrões globais de comportamento.

Existem diversos conjuntos de dados que podem ser considerados para este efeito mas os mais relevantes são os que dizem respeito ao número de alunos inscritos, avaliados e aprovados nas diversas unidades curriculares.

Uma simples representação tabular pode não facilitar a análise destes dados, ao contrário do uso de técnicas de Visualização de Informação.

A Visualização de Informação tem como principal objetivo auxiliar as pessoas a compreender os dados representados, aproveitando a capacidade da visão humana para detetar padrões, identificar padrões e anomalias [1].

1.2 Objetivos

Este trabalho de dissertação tem como objetivo o desenvolvimento de uma plataforma web que permita aos diretores de curso fazer uma análise do desempenho dos alunos nas diversas unidades curriculares.

Essa plataforma permite não só aceder ao número de alunos inscritos, avaliados e aprovados às unidades curriculares mas também permite aceder a um conjunto de indicadores, como por exemplo às taxas de abandono e de aprovação nas unidades curriculares selecionadas.

A plataforma deverá apresentar estes números e indicadores através de representações visuais que irão facilitar a análise dos dados por parte do utilizador.

Para isso será necessário proceder a uma recolha de dados e tratá-los, criar um repositório para os armazenar, estudar as possíveis tecnologias a ser usadas no desenvolvimento da plataforma e implementá-la e, finalmente, avaliá-la.

1.3 Estrutura da dissertação

No capítulo 2 é apresentada uma introdução aos sistemas de suporte à decisão e à visualização de informação, descrevendo o que são e que ferramentas e tecnologias podem ser utilizados na sua implementação. É também apresentado um trabalho já existente com objetivos semelhantes aos deste trabalho de dissertação.

O Capítulo 3 apresenta os requisitos deste projeto, as tecnologias utilizadas, a estrutura adotada na sua implementação e é feita ainda uma apresentação da plataforma resultante.

No Capítulo 4 é apresentado o resultado das avaliações feitas à plataforma desenvolvida.

Por fim, no Capítulo 5, são apresentadas as conclusões, assim como sugestões para trabalho futuro.

2. Estado da Arte

Para compreender a eficácia dos processos de ensino-aprendizagem, identificar pontos fracos ou definir áreas de atuação tendo em vista a melhoria dos processos educativos, importa ter acesso a informação relevante, atualizada, fidedigna e facilmente interpretável. No entanto, apesar de hoje em dia já existirem ferramentas adequadas ao desenvolvimento de sistemas de informação capazes de facilitar estes processos, ainda existe um longo caminho a percorrer na área da educação.

Este capítulo faz uma apresentação de metodologias e ferramentas vocacionadas para o desenvolvimento de sistemas de apoio à decisão e visualização de informação. No primeiro caso, será dada ênfase aos processos de modelação, organização e processamento analítico de dados, e no segundo caso será dada ênfase à representação gráfica de informação educacional. No final, também é apresentado um trabalho relacionado com a visualização e interpretação de dados ao nível do ensino superior.

2.1 Sistemas de Suporte à Decisão

Sistemas de suporte à decisão é uma expressão utilizada na área de sistemas de informação para designar um conjunto de ferramentas interativas que permitem auxiliar os utilizadores em processos de tomada de decisões. Estes sistemas normalmente permitem selecionar, agregar, visualizar e comparar dados de forma interativa, ou mesmo fazer a previsão de possíveis resultados no futuro tendo em conta experiências anteriores. Os dados são normalmente armazenados em *Data Warehouses* que podem ser vistos como bases de dados multidimensionais que podem ser utilizadas em processos analíticos, como o OLAP, *data mining* ou inteligência artificial [2] [3].

2.1.1 Metodologias de Desenvolvimento

A necessidade de agregar grandes volumes de dados provenientes de diversos sistemas levou a que fosse desenvolvido um novo género de Sistema de Informação designado de *Data Warehouse*. Este Sistema de Informação é construído com o intuito de apoiar o processo de tomada de decisão na organização [4].

Um *Data Warehouse* é um grande repositório de dados integrados obtidos de várias fontes com o único objetivo de análise dos dados que influenciam no processo de tomar decisões.

É construído para permitir: uma vista integrada e completa de toda a organização; acesso aos dados históricos da organização; ter uma fonte de dados dentro da organização e facilitar os processos de tomada de decisão, sem sobrecarregar os sistemas operacionais [5].

A Figura 1 apresenta o esquema do armazenamento de dados descrito.

Como se pode verificar, os dados não tratados são transformados em dados informais que são armazenados, sendo depois usados para suportar outras funções como a planificação e a previsão [6], [7], [8].

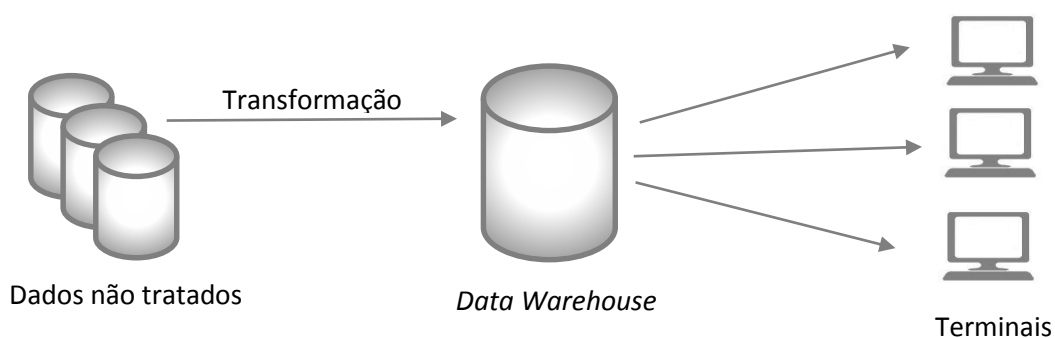


Figura 1 – Esquema de armazenamento de dados num Data Warehouse

Existe um grande conjunto de ferramentas que podem ser utilizadas para implementar um *Data Warehouse*, mas a modelação deste está geralmente baseada numa de duas metodologias: Inmon ou Kimball.

Metodologia de Kimball

Ralph Kimball define um *Data Warehouse* como:

“... a copy of transaction data specifically structured for query and analysis.” [9]

A metodologia de Kimball, que se encontra representada na Figura 2, baseia-se num ciclo que inicia com o planeamento do projeto. Nesta fase são estabelecidos os objetivos, são adquiridos os recursos necessários e é iniciado o projeto.

O ciclo continua com a definição dos requisitos do projeto. Uma vez que os requisitos impostos pelos utilizadores têm um grande impacto nas decisões relativas ao desenho e à implementação de um projeto, esta fase é seguida por três sequências de tarefas paralelas.

A primeira tarefa da sequência lida com a tecnologia do projeto e é composta pelo desenho técnico da arquitetura, que estabelece uma framework que permite integrar várias tecnologias, e pela seleção do produto, que é feita apenas depois de terem sido identificadas no desenho da arquitetura todos os recursos necessários.

A segunda tarefa da sequência foca-se nos dados e começa pela tradução dos requisitos num modelo dimensional que, depois, é transformado numa estrutura física, que é baseada em estratégias para melhoramento do desempenho, como agregações, indexação e particionamento. Por fim é feito o desenho e o desenvolvimento dos processos de extração dos dados (ETL – Extract – Transformation - Load).

A última tarefa é composta pelo desenho e o desenvolvimento das aplicações analíticas.

Após estas tarefas paralelas é implementado o projeto.

Por último, é necessário assegurar que o projeto permanece sem problemas e, por isso, existe uma fase de manutenção que se prolonga até ao final do projeto.

Em paralelo existe uma fase de crescimento que se foca no futuro do projeto. Este ciclo de vida do modelo de Kimball é aplicado desde o início sobre qualquer subprojecto que surja desta fase [10] [11].

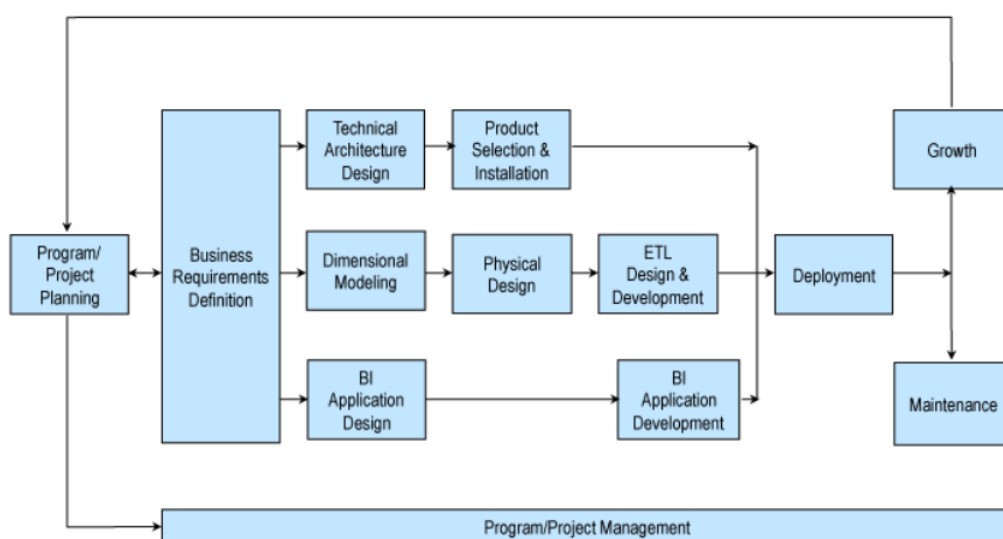


Figura 2 – Metodologia de Kimball [9]

Metodologia de Inmon

Bill Inmon define um *Data Warehouse* como:

“... a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions.” [12]

A metodologia de Inmon, representada na Figura 3, é constituída por um conjunto de fases.

A primeira fase consiste na análise do modelo de dados (DSS1) seguida da fase onde é estimado o volume de informação planeado para o data warehouse para que possam ser tomadas as devidas precauções (DSS2). Em paralelo com esta tarefa é preparada a arquitetura tecnológica de suporte (DSS3) e o ambiente técnico (DSS4). De seguida é feita a escolha do tema a ser implementado (DSS5), é feita a análise das fontes e identificada a origem dos dados a transferir para o *Data Warehouse* (DSS7).

A fase seguinte consiste na elaboração do desenho da base de dados para o *Data Warehouse* (DSS6).

As últimas três fases deste modelo consistem na definição das ferramentas que irão ser responsáveis pelas operações, extrações, transformação e integração dos dados provenientes das diversas fontes de informação (DSS8), o estudo e o desenvolvimento dessas mesmas ferramentas (DSS9) e o carregamento dos dados para o *Data Warehouse* utilizando as ferramentas desenvolvidas (DSS10).

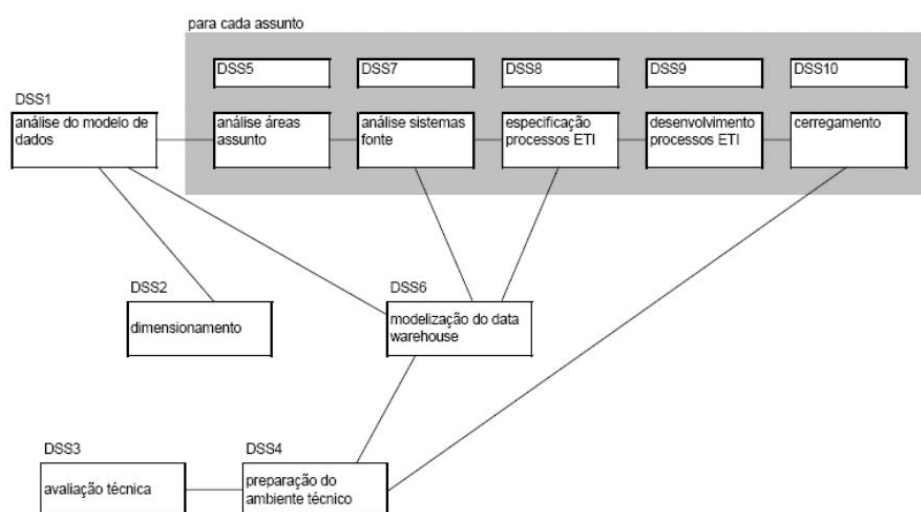


Figura 3- Metodologia de Inmon. Adaptado de [13]

Comparando as duas metodologias de desenvolvimento apresentadas chega-se à conclusão que os dois modelos são opostos e defendem procedimentos diferentes. No entanto, existem algumas diferenças que se destacam mais:

- O modelo de Inmon baseia-se num estilo mais tradicional na construção da base de dados enquanto o modelo de Kimball centra-se num estilo incremental onde cada *Data Mart* vai sendo integrado à medida que o projeto vai evoluindo.
- O modelo de Inmon tem uma arquitetura *top-down* enquanto o modelo de Kimball tem uma arquitetura *bottom-up*
- A arquitetura utilizada no modelo de Inmon é mais centralizada nos dados, o que a torna mais morosa e dispendiosa e inviabiliza o envolvimento dos utilizadores no projeto. Por outro lado, a arquitetura utilizada no modelo de Kimball é mais centralizada nos processos individuais, o que permite antecipar a entrega de resultados. Garante ainda um maior envolvimento dos utilizadores e permite fasear os custos [13].

2.1.2 Modelos de Dados Multidimensionais

No *Data Warehousing* são utilizados modelos de dados multidimensionais.

Os modelos de dados multidimensionais são modelos desenhados expressamente para suportar análise de dados. Estes modelos são compostos por várias estruturas de dados que servem para capturar os dados e analisá-los, designadas cubos. Cada célula de um cubo, resultante da intersecção de várias dimensões, é denominada de facto.

Um facto é definindo como sendo um objeto que representa o sujeito que se pretende analisar e tem associadas a si várias medidas que o caracterizam e uma granularidade que define o nível de detalhe. Estas medidas podem ser simplesmente valores numéricos ou fórmulas (normalmente funções simples, como por exemplo, somas).

2.1.3 OLAP – Processamento Analítico de Dados

Na análise de *Data Warehouses* é utilizado o processo OLAP.

OLAP (*On-Line Analytical Processing*) é um sistema de manipulação e análise de um grande volume de dados sob várias perspetivas que possui ferramentas que podem ser usadas em sistemas de suporte à decisão. Alguns dos mecanismos utilizados por este tipo de sistemas são as agregações e pré-cálculos, mas também é possível realizar consultas *ad hoc* para suportar sistemas de suporte à decisão.

Existem várias categorias OLAP: O MOLAP, o ROLAP e o HOLAP.

No MOLAP (*Multidimensional OLAP*) os dados são modelados e armazenados numa base de dados multidimensional e as suas ferramentas são implementadas por DBMS (*Database Management Systems*) especializados e capazes de suportar dados multidimensionais diretamente.

No ROLAP (*Relational OLAP*) os dados são armazenados numa base de dados relacional e um servidor ROLAP cria uma vista multidimensional para o utilizador. As suas ferramentas são menos complexas mas também menos eficientes.

O HOLAP (*Hybrid OLAP*) é uma combinação do MOLAP com o ROLAP. Os dados que não são atualizados frequentemente são armazenados numa base de dados multidimensional enquanto os dados que são atualizados frequentemente são armazenados numa base de dados relacional.

2.1.4 Ferramentas e Tecnologias

Existem várias ferramentas que podem ser utilizadas neste trabalho de dissertação com o objetivo de facilitar o desenho e a manutenção de um *Data Warehouse*, assim como a integração e a migração dos dados a partir de várias fontes.

Nesta secção irão ser analisadas algumas dessas ferramentas.

Pentaho

O *Pentaho* é uma plataforma com componentes que permitem desenvolver soluções completas para os problemas de BI (*Business Intelligence*).

Esta plataforma permite ao utilizador fazer relatórios, análise de fluxo de trabalho, painéis e *data mining*.

Inclui componentes e relatórios para analisar o desempenho dos processos que são executados. As operações da plataforma são especificadas nas definições de processos e nos documentos de ação que descrevem todas as atividades. Estes processos e operações definem a solução para um problema de BI, que pode ser facilmente integrada em processos que são externos à plataforma.

Esta ferramenta necessita de uma licença paga, contudo possui um período de experimentação de 30 dias [14] [15].

IBM Analytical Decision Management:

O *IBM Analytical Decision Management* é uma ferramenta de apoio à decisão orientada para empresas. Permite às empresas tomarem as melhores decisões de forma a atingirem os objetivos pretendidos e evitarem riscos de investimento desnecessários.

Esta aplicação possui ferramentas de previsão, técnicas matemáticas para otimizar decisões transicionais, modelos de previsão combinados e integrados, regras e lógica de decisões para obter as ações recomendadas. Disponibiliza simulações para acomodar mudanças nas condições com base no volume, variedade e velocidade de dados de entrada e possui uma interface flexível e intuitiva para apoiar o desenvolvimento e implementação de configurações e conteúdo. Permite ainda a integração com outras soluções de *software* da IBM.

O *IBM Analytical Decision Management* é uma ferramenta que necessita de licença paga [16] [17].

MicroStrategy Analytics:

O *MicroStrategy Analytics* é uma plataforma desenvolvida utilizando apenas um servidor de forma a permitir segurança e administração centralizadas, assim como a integração de vários produtos. Fornece a opção de criar um único modelo de dados ou combinar dados de diferentes fontes em tempo real. Isto proporciona uma grande flexibilidade na análise dos dados.

Esta plataforma baseia-se numa arquitetura ROLAP e cria automaticamente, em memória, um conjunto ideal de cubos com base em padrões encontrados na utilização da plataforma por parte de vários utilizadores. Este processo reduz o trabalho da administração e melhora os tempos de resposta.

O *MicroStrategy* permite a definição de uma única visualização com atributos / métricas de vários conjuntos de dados (relatórios / cubos).

Apesar de ser uma ferramenta que necessita de uma licença paga, disponibiliza uma versão grátis com funcionalidades limitadas [18] [19].

2.2 Visualização de Informação

Visualização de informação é um processo que produz representações visuais e interativas com o objetivo de facilitar ao utilizador a compreensão dos dados.

Chaomei Chen completa esta definição dizendo:

“Information visualization is concerned with the design, development and application of computer generated interactive graphical representations of information.” [20]

Este processo é constituído por várias fases que irão ser descritas nesta secção, assim como as diversas ferramentas que podem ser utilizadas na criação de uma visualização.

2.2.1 Processo de visualização de informação

Na Figura 4 é apresentada uma representação visual do processo de visualização de informação.

A primeira fase deste processo consiste na **transformação e análise dos dados**, isto é, os dados obtidos são analisados e transformados de forma a facilitar as fases seguintes do processo.

A segunda fase consiste na **filtragem** dos dados, ou seja, consiste na seleção de um subconjunto de dados. Esta fase permite evitar problemas de desempenho no caso de haver um grande conjunto de dados.

Na fase seguinte é feito o **mapeamento** dos dados filtrados. Nesta fase, os dados são representados utilizando primitivas, como pontos e segmentos de reta, e são associados atributos, como cor, posição ou tamanho.

A fase seguinte é a fase da **renderização**, onde o mapeamento da fase anterior é transformado em representações visuais.

Por fim, é criado um conjunto de **controles de interação** entre o utilizador e as representações. Estes controlos permitem ao utilizador analisar os dados representados e explorá-los de forma a melhorar a sua compreensão.

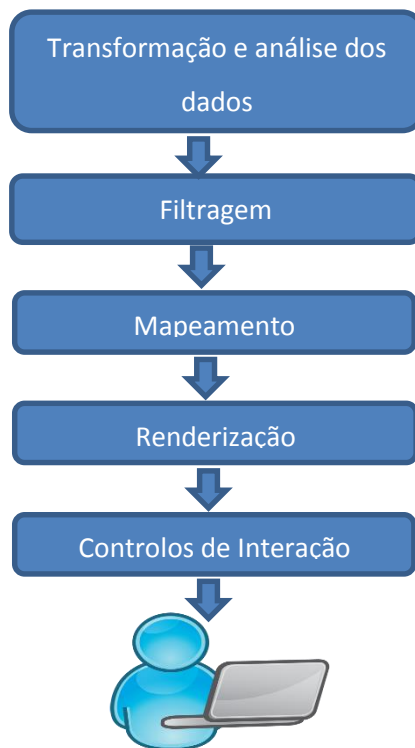


Figura 4 – Processo de visualização de informação [21]

2.2.2 Ferramentas e Tecnologias

Para a Visualização de Informação existem diversas ferramentas para desenvolvimento de aplicações desktop assim como bibliotecas direcionadas para desenvolvimento web.

Para este projeto será mais vantajoso utilizar uma biblioteca para desenvolvimento web, uma vez que se trata de uma plataforma web. Algumas das bibliotecas consideradas foram Google Charts, Highcharts JS, ChartJS e d3.js.

Google Charts

O Google Charts é uma biblioteca que possui uma galeria com diversas visualizações que podem ser personalizados através das várias opções disponibilizadas, o que facilita a criação de uma visualização.

O Google Charts tem ainda as vantagens de ser uma ferramenta de licença gratuita e compatível com a maioria dos *browsers*.

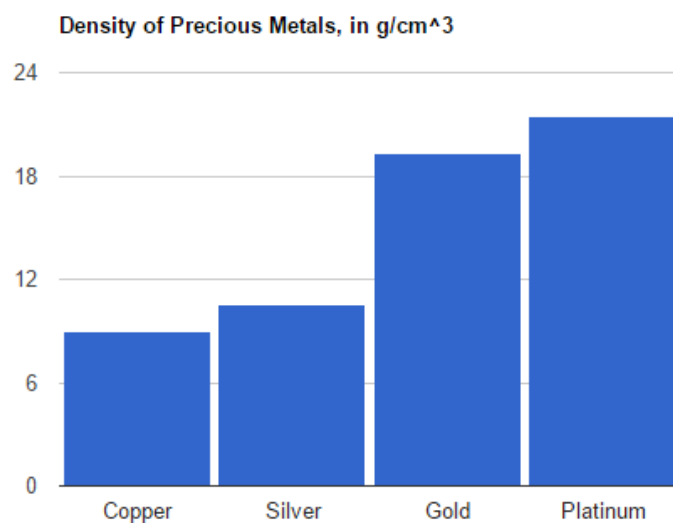


Figura 5 – Gráfico de barras construído utilizando a biblioteca Google Charts [22]

Highcharts JS

O Highcharts JS é outra biblioteca bastante popular que permite criar uma visualização com diversas animações.

À semelhança do Google Charts, possui uma vasta galeria de visualizações que podem ser personalizados e é uma ferramenta de licença gratuita.

Uma das maiores vantagens desta ferramenta é a compatibilidade com *browsers* mais antigos.

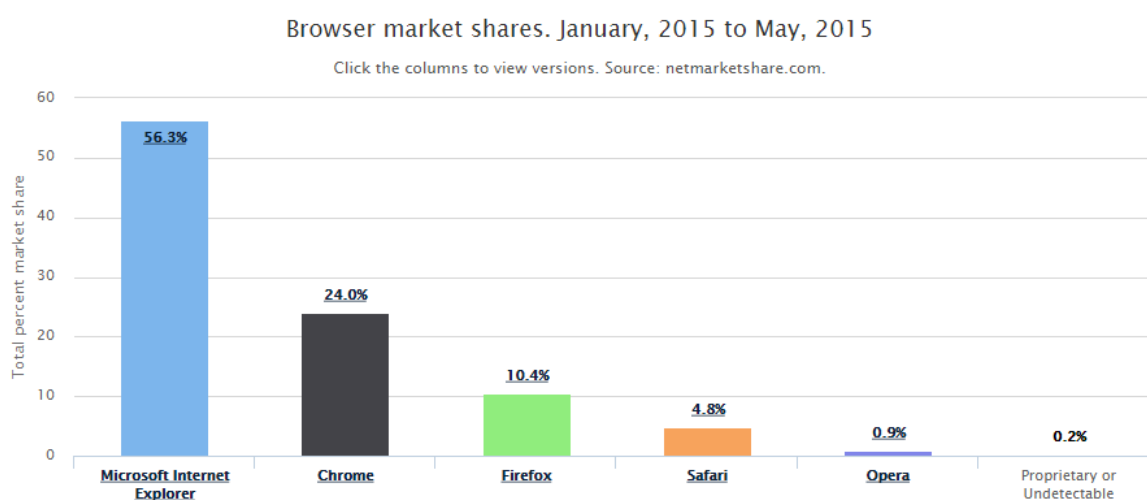


Figura 6 – Gráfico de barras construído utilizando a biblioteca Highcharts JS [23]

ChartJS

O ChartJS é uma biblioteca que permite a criação de 6 tipos de visualizações diferentes com um *design* atrativo, é compatível com todos os *browsers* recentes e possui licença gratuita.

A grande vantagem desta ferramenta é o fato de todas as visualizações serem responsivas por omissão.

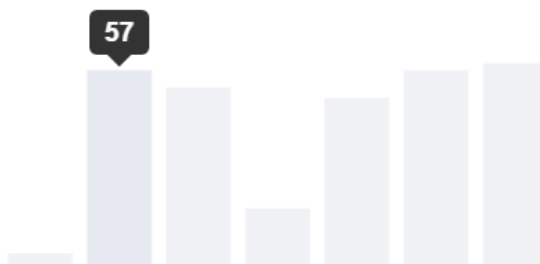


Figura 7 – Gráfico construído utilizando a biblioteca Charts JS [24]

d3.js

O d3.js é uma ferramenta que permite a criação de visualizações dinâmicas, possui diversas opções de personalização e a sua licença é gratuita. Esta ferramenta não possui visualizações pré construídas mas existe uma lista bastante extensa de visualizações construídas utilizando o d3.js que podem servir de exemplo.

Pode ter alguns problemas de compatibilidade com *browsers* mais antigos mas existem vários *plugins* que podem ser utilizados para resolver esses problemas, como por exemplo o *aight plugin*.

Tem como principal vantagem o facto de suportar uma grande quantidade de dados.

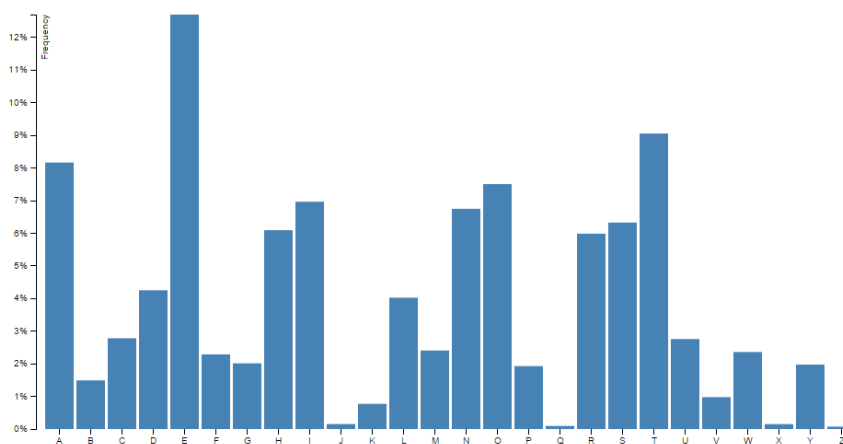


Figura 8 – Gráfico construído utilizando a biblioteca d3.js [25]

2.3 Sistemas de apoio à decisão na educação e visualização de dados educacionais

Existem já algumas plataformas desenvolvidas para visualização de dados educacionais. Um exemplo destas plataformas é o EduVis.

O EduVis é uma plataforma que foi desenvolvida no âmbito do projeto EduCare por Sandra Gama, Vilma Jordão e Daniel Gonçalves, no Instituto Superior Técnico da Universidade de Lisboa em 2014.

“Compreender a eficácia de um programa de estudos ou de uma disciplina implica perceber como os alunos atingem o sucesso” [26]

Tem como objetivo apresentar relações entre as várias unidades curriculares através de representações gráficas.

Estas relações são apresentadas utilizando representações multicamada das unidades curriculares interrelacionadas através de conectores visuais. Cada nível da representação corresponde a um semestre curricular e as unidades curriculares são representadas por círculos. Os círculos são compostos por uma parte verde e uma parte vermelha que correspondem ao número de alunos com sucesso e ao número de alunos sem sucesso, respetivamente.

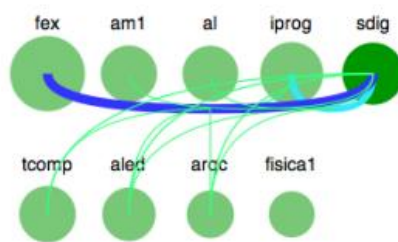


Figura 9 – Representação multicamada[26]

São também utilizadas representações multimatriciais onde são apresentadas as unidades curriculares, sendo os respetivos padrões representados por mapas de cor. Cada matriz corresponde a uma unidade curricular e é dividida em dois triângulos: o triângulo superior representa o número de alunos com sucesso e o triângulo inferior representa o número de alunos sem sucesso.

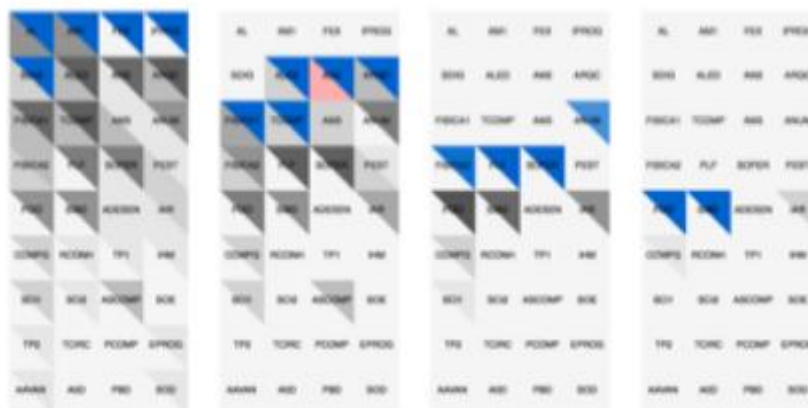


Figura 10 – Representação multimatricial [26]

Através destes dois tipos de representação são analisados os padrões e encontradas relações entre as várias unidades curriculares.

2.4 Sumário

Após a análise das ferramentas de sistemas de apoio à decisão concluiu-se que a ferramenta mais viável seria a MicroStrategy Analytics por disponibilizar uma versão gratuita apesar de possuir algumas limitações. Uma delas é o facto de não permitir apresentar hierarquias nas representações gráficas, sendo apenas possível visualizar as hierarquias pretendidas adicionando ou retirando dimensões dos parâmetros de visualização. Contudo, optou-se por não utilizar nenhuma destas ferramentas, uma vez que nenhuma traz vantagens para este projeto.

Apesar de todas as ferramentas de visualização de informação serem semelhantes concluiu-se que a d3.js seria a mais vantajosa para o desenvolvimento deste projeto uma vez que é a única que suporta uma grande quantidade de dados.

3. Desenvolvimento

Neste capítulo serão apresentados e descritos os requisitos funcionais do sistema, os atores envolvidos e respetivos casos de uso, a forma como foi organizado e desenvolvido este trabalho de dissertação e a base de dados associada e, finalmente, será apresentada a plataforma resultante.

O desenvolvimento do trabalho foi feito com base na metodologia de Kimball descrita no capítulo 2 desta dissertação.

3.1 Requisitos funcionais

Inicialmente foram definidos os requisitos do sistema considerando os objetivos do trabalho de dissertação proposto. Assim, o sistema deve permitir:

- Agregar de forma organizada uma grande quantidade de dados de diversas fontes
- Permitir ao utilizador selecionar alguns critérios para seleção de dados
- Filtrar um subconjunto de dados tendo em conta os critérios definidos pelo utilizador
- Representar os dados selecionados através de gráficos de barras ou de linhas, consoante o tipo de dados que se pretende visualizar.

3.2 Atores

O sistema tem como público-alvo professores e diretores de curso que pretendam analisar o desempenho dos alunos nas diversas unidades curriculares.

Considerando os requisitos especificados anteriormente foram identificados os seguintes atores:

1. Professor / Diretor de Curso – é o utilizador da plataforma que pretende visualizar os dados
2. Sistema – entidade computacional que trata os dados e cria as visualizações selecionadas pelo utilizador.

3.3 Modelo de casos de uso

Atendendo aos requisitos e os atores identificados, foi elaborado modelo de casos de uso representado na Figura 11.

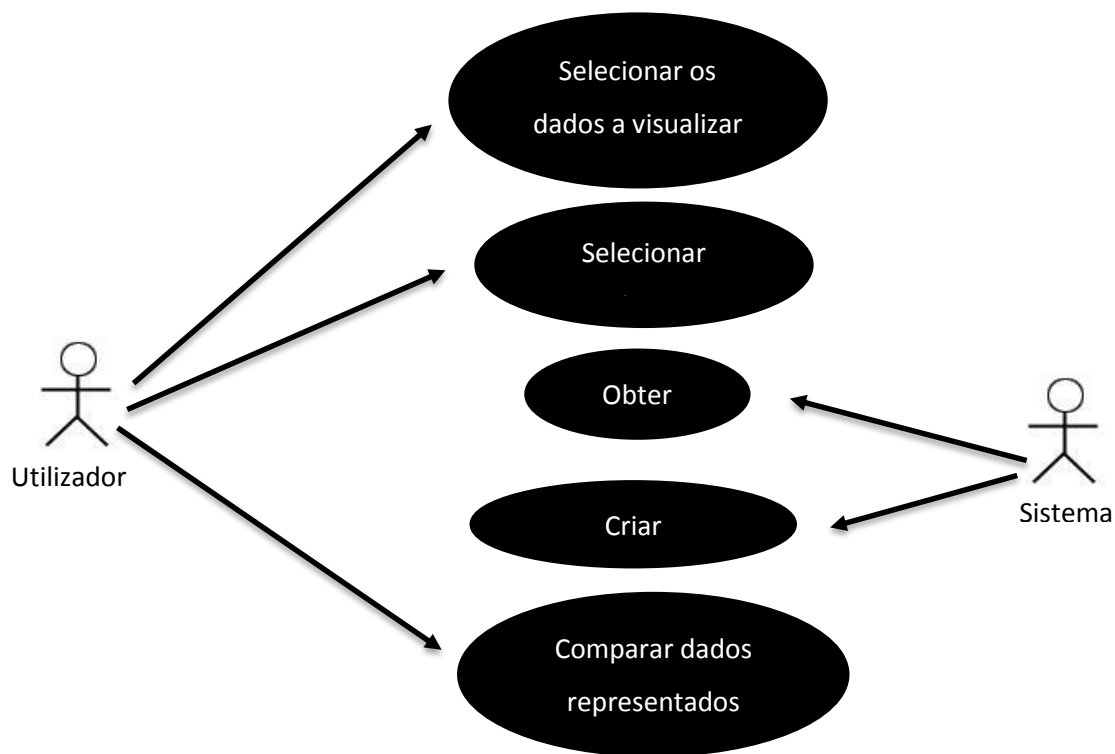


Figura 11 – Modelo de casos de uso

3.4 Descrição dos casos de uso

Os casos de uso do utilizador, representados na Figura 12, são os seguintes:

1. Selecionar os dados a visualizar – inicialmente, o utilizador escolhe que tipo de dados pretende visualizar entre a visualização de taxas de aprovação e a visualização de número de alunos inscritos/avaliados/aprovados.
2. Selecionar os critérios de seleção – dependendo do tipo de dados pretendido, o utilizador seleciona um ou mais critérios de seleção. Estes critérios de seleção podem ser o curso, unidade curricular, ano letivo, ano curricular, semestre curricular e área científica. No caso de se pretender visualizar o número de alunos inscritos/avaliados/aprovados, o utilizador poderá visualizar apenas um deles ou os três em simultâneo, dependendo da comparação que pretende efetuar.
3. Comparar dados representados – O utilizador analisa a visualização gerada pelo sistema e compara os dados representados, consoante o que pretende.

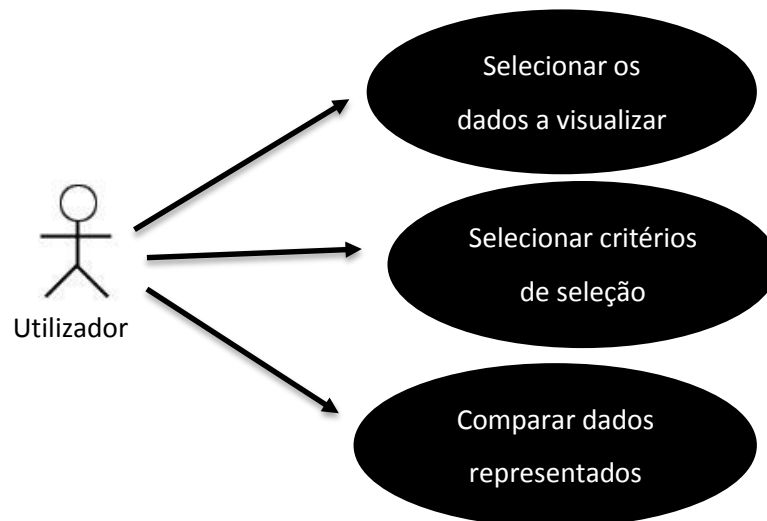


Figura 12 – Modelo de casos de uso do utilizador

Os casos de uso do sistema, representados na Figura 13, são os seguintes:

1. Obter dados – Tendo em conta os critérios selecionados pelo utilizador, o sistema obtém os dados associados.
2. Criar representação – O sistema representa os dados obtidos utilizando o tipo de visualização mais adequado.

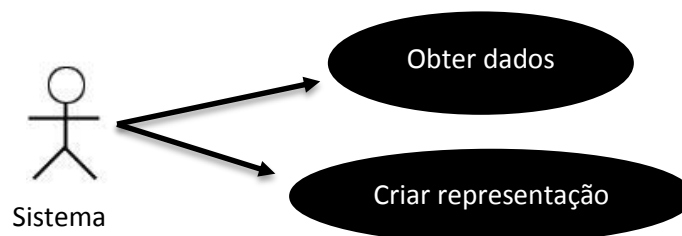


Figura 13 – Modelo de casos de uso do sistema

3.5 Estrutura do sistema

Tendo em conta os requisitos definidos e os casos de uso de cada ator descritos anteriormente foi desenhada uma plataforma utilizando a estrutura MCV (Model-View-Controller).

Esta estrutura, representada na Figura 14, consiste na separação entre a representação da informação e interação do utilizador com ela e é composta por 3 partes: o modelo, a visão e o controlador.

O modelo (Model) consiste na interação com os dados, como por exemplo, a consulta da base de dados através de *queries*.

A visão (View) consiste na representação dos dados através de tabelas, gráficos, etc.

O controlador (Controller) faz a ligação entre o modelo e a visão, recebendo as informações do modelo e convertendo-as numa resposta para a visão, ou vice-versa.

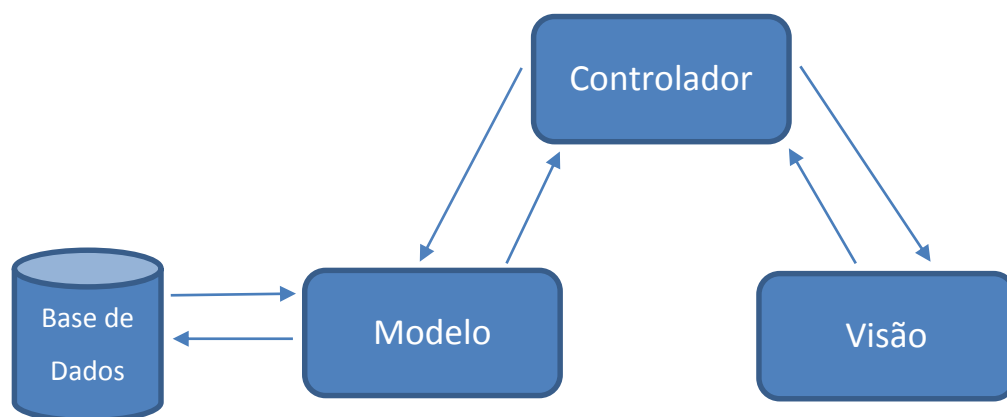


Figura 14 – Esquema da estrutura MVC

Considerando a estrutura acima descrita, foi desenhado um diagrama composto por 2 componentes principais: a aplicação do cliente e o servidor.

A aplicação do cliente é constituída pela visão, ou seja, o ambiente gráfico que será utilizado para interagir com a plataforma tanto para a escolha dos critérios de seleção como para a apresentação das visualizações geradas pelo sistema. A visão foi desenvolvida utilizando a linguagem HTML com a biblioteca bootstrap para o desenvolvimento ambiente gráfico e as bibliotecas de javascript JQuery e d3.js para o tratamento dos critérios de seleção e a criação das visualizações.

O servidor é composto pelo controlador, pelo modelo e pela base de dados. O modelo e o controlador foram desenvolvidos utilizando a linguagem PHP e a base de dados foi criada utilizando SQL Server.

Para a população da base de dados foi utilizado o processo ETL (Extract – Transformation – Load) utilizando a ferramenta de importar dados disponibilizada pelo SQL Server. Este processo consiste na extração dos dados disponibilizados, na sua transformação para a estrutura da base de dados pretendida e, finalmente, na sua transferência para a base de dados.

Para estabelecer comunicação entre o cliente e o servidor são enviados pedidos através do método POST a partir da visão para o controlador.

O controlador recebe estes pedidos, valida os parâmetros enviados pela visão (se tiver sido enviado algum) e chama a função pretendida do modelo. A função do modelo que foi invocada executa a *query* a si associada, utilizando o *plugin* PDO (PHP Data Objects) do PHP, e devolve os resultados ao controlador. O controlador valida a resposta recebida do modelo, converte-a no formato *json* utilizando a função *json_encode* do PHP e envia-a para a visão. A visão transforma a resposta recebida no formato necessário para a criação da visualização pretendida.

Atendendo ao descrito anteriormente foi desenvolvido o diagrama de componentes representado na Figura 15.

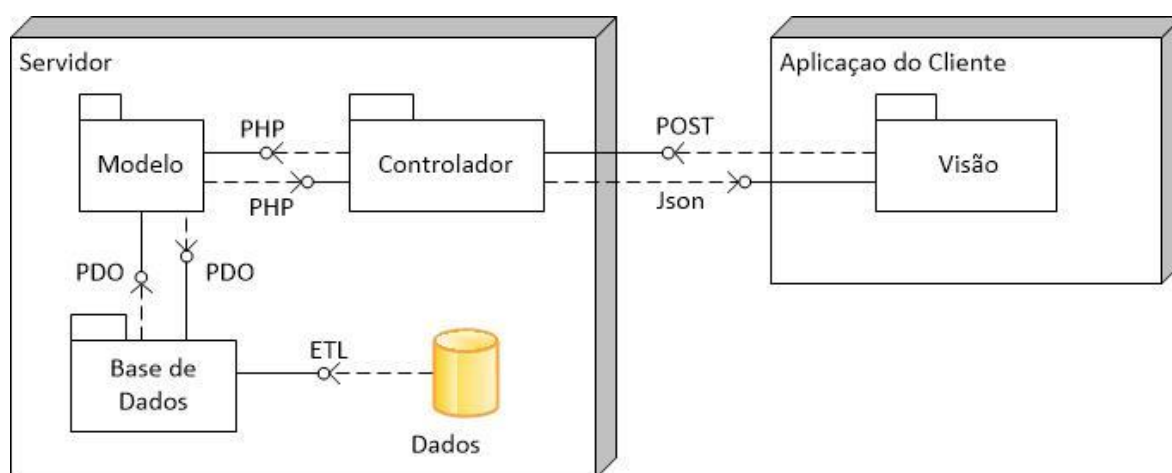


Figura 15 – Diagrama de Componentes

3.6 Organização dos Dados

Para que os dados fossem organizados de forma a serem facilmente consultados foi criada uma base de dados multidimensional baseada no sistema ROLAP, descrito no capítulo 2 desta dissertação.

Para facilitar a modelação da base de dados foram formuladas algumas perguntas que deverão poder ser respondidas no final. Essas perguntas foram organizadas no seguinte esquema:

- Taxa de aprovação dos alunos inscritos numa dada disciplina
 - Para um curso apenas
 - Para um ano
 - Para os últimos anos
 - Diferenças entre os vários cursos que tenham a mesma disciplina
 - Para um ano
 - Para os últimos anos

- Taxa de aprovação dos alunos avaliados numa dada disciplina
 - Para um curso apenas
 - Para um ano
 - Para os últimos anos
 - Diferenças entre os vários cursos que tenham a mesma disciplina
 - Para um ano
 - Para os últimos anos
- Taxa de aprovação num semestre curricular
 - Para um curso apenas
 - Para um ano
 - Para os últimos anos
 - Diferenças entre os vários cursos do departamento
 - Para um ano
 - Para os últimos anos
- Taxa de aprovação de uma dada área científica
 - Para um curso apenas
 - Para um ano
 - Para os últimos anos
 - Diferenças entre os vários cursos do departamento
 - Para um ano
 - Para os últimos anos
- Taxa de aprovação de uma “sequência” de disciplinas
 - Para um curso apenas
 - Para um ano
 - Para os últimos anos
 - Diferenças entre os vários cursos do departamento
 - Para um ano
 - Para os últimos anos
- Diferença entre alunos inscritos e avaliados em cada disciplina
 - Para um curso apenas
 - Para um ano
 - Para os últimos anos
 - Diferenças entre os vários cursos do departamento
 - Para um ano
 - Para os últimos anos

Com o objetivo de responder às perguntas formuladas foram criadas duas tabelas de factos que têm associadas as mesmas medidas (número de alunos inscritos, número de alunos avaliados e número de alunos aprovados).

Para caracterizar estas medidas, a tabela de factos **fact1** utiliza as dimensões **UC** (que representa uma unidade curricular), **epoca** (que representa a época em que os resultados foram registados), **anoLectivo** (que representa o ano letivo em questão) e **semestre** (que representa o semestre em questão).

A tabela de factos **fact2** utiliza as dimensões **curso** (que representa um curso), **UC** (que representa uma unidade curricular), **epoca** (que representa a época em que os resultados foram registados), **anoLectivo** (que representa o ano letivo em questão) e **semestre** (que representa o semestre em questão). Foram ainda criadas outras duas tabelas de factos, a tabela **fact3** e a tabela **fact4**.

A tabela de factos **fact3** representa o número de alunos inscritos por ano curricular num dado curso e num certo ano letivo. Tem associada a si apenas uma medida **nr_inscritos** (que representa o número de alunos inscritos) e utiliza as dimensões **anoLectivo** (que representa o ano letivo em questão), **curso** (que representa um curso) e **anoCurricular** (que representa ano curricular a considerar).

A tabela de factos **fact4** representa a proveniência de inscrições e razões de abandono de um dado curso num certo ano letivo. Utiliza as dimensões **anoLectivo** (que representa o ano letivo em questão) e **curso** (que representa um curso) e tem associadas a si as medidas **inConcNacAcess** (que representa o número de alunos que entraram no curso por concurso nacional de acesso ao ensino superior), **inConcIntAcess** (que representa o número de alunos que entraram no curso por concurso internacional de acesso ao ensino superior), **inTitCurSup** (que representa o número de alunos que entraram no curso por candidatura especial de titular de curso superior), **inTitCurPosSec** (que representa o número de alunos que entraram no curso por candidatura especial de titular de curso pós-secundário), **inReingresso** (que representa o número de alunos que entraram no curso por reingresso), **inMudCurso** (que representa o número de alunos que entraram no curso por mudança de curso), **inTransferencia** (que representa o número de alunos que entraram no curso por transferência), **in2CicloEdital** (que representa o número de alunos que entraram no 2º ciclo do curso por edital), **in2CicloCoeCient** (que representa o número de alunos que entraram no 2º ciclo do curso por coerência científica), **outTransferencia** (que representa o número de alunos que saíram do curso por transferência), **outMudCurso** (que representa o número de alunos que saíram do curso por mudança de curso), **outConclusao** (que representa o número de alunos que saíram do curso por terem concluído), **outDesistencia** (que representa o número de alunos que saíram do curso por terem desistido) e **outAbandono** (que representa o número de alunos que saíram do curso por terem abandonado).

Para além das tabelas de factos foram criadas as tabelas para as dimensões necessárias.

Na Figura 16 está representado o modelo resultante.

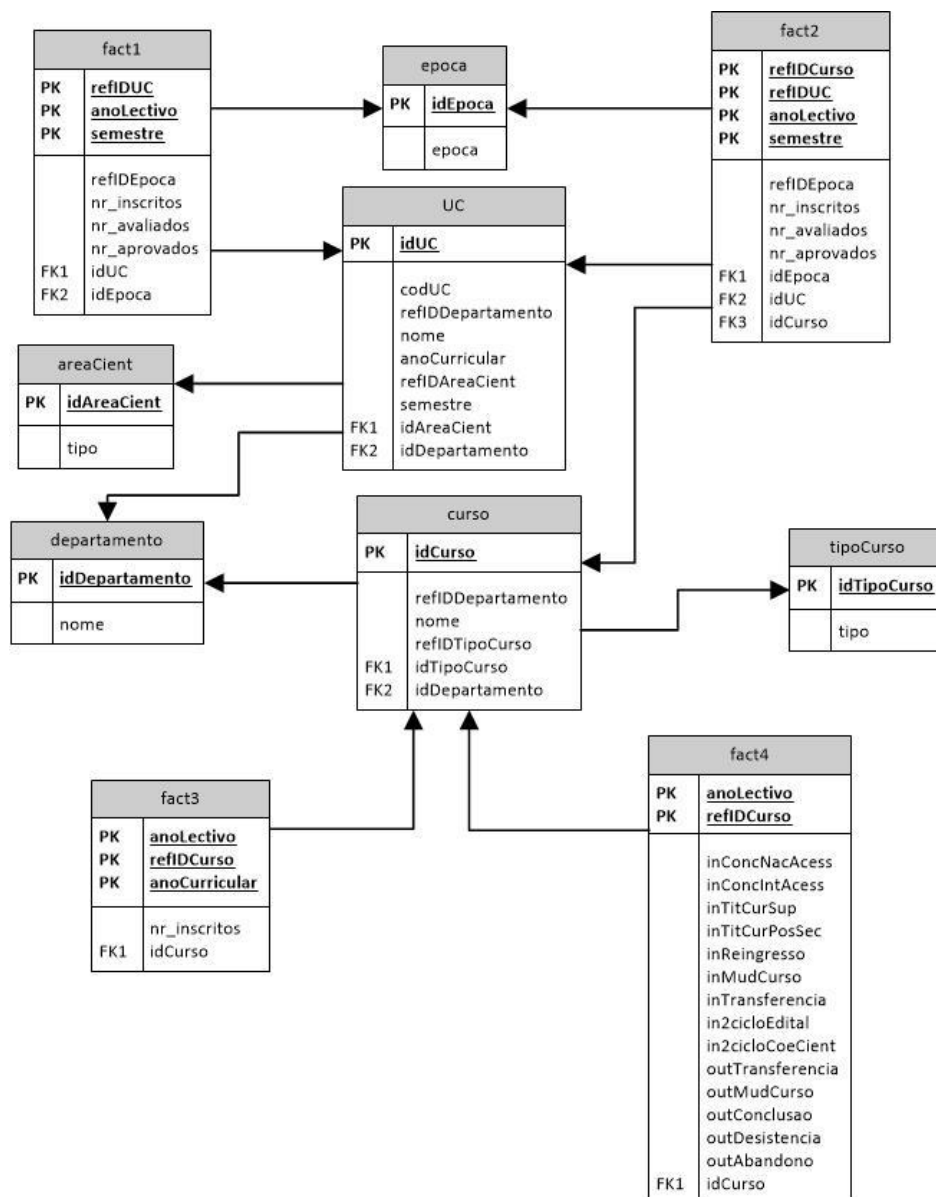


Figura 16 – Modelo da Base de Dados Multidimensional

Para a população da base de dados foi necessária a criação de um ficheiro Excel com todos os dados fornecidos agregados numa estrutura que permitisse a importação direta dos dados para o SQL Server, uma vez que os dados foram disponibilizados em diversos ficheiros Excel cada um com a sua estrutura.

A Figura 17 apresenta um excerto do ficheiro Excel criado.

	A	B	C	D	E	F	G	H	I	J	K
1	AnoLectivo	Dep	CodCurso	NomeCurso	CodDiscip	idUC	Semestre	Inscritos	Avaliados	Aprovados	Epoca
2	2009	dlc	8204	Engenharia Electrónica e Telecomunicações	45983	125	0	5	5	5	4
3	2009	dlc	8204	Engenharia Electrónica e Telecomunicações	45985	128	0	4	4	4	4
4	2009	dlc	8204	Engenharia Electrónica e Telecomunicações	45986	130	0	1	1	1	4
5	2009	dlc	8204	Engenharia Electrónica e Telecomunicações	45996	136	0	5	5	5	4
6	2009	dlc	8204	Engenharia Electrónica e Telecomunicações	45998	138	0	3	3	3	4
7	2009	dlc	8204	Engenharia Electrónica e Telecomunicações	46026	150	0	1	1	1	4
8	2009	deti	8204	Engenharia Electrónica e Telecomunicações	49952	216	0	81	81	81	4
9	2010	dlc	8204	Engenharia Electrónica e Telecomunicações	45983	125	0	3	3	3	4
10	2010	dlc	8204	Engenharia Electrónica e Telecomunicações	45985	128	0	1	1	1	4
11	2010	dlc	8204	Engenharia Electrónica e Telecomunicações	45996	136	0	1	1	1	4
12	2010	dlc	8204	Engenharia Electrónica e Telecomunicações	46026	150	0	1	1	1	4
13	2010	deti	8204	Engenharia Electrónica e Telecomunicações	49952	216	0	62	62	62	4
14	2011	dlc	8204	Engenharia Electrónica e Telecomunicações	45983	125	0	2	2	2	4
15	2011	dlc	8204	Engenharia Electrónica e Telecomunicações	45985	128	0	1	1	1	4
16	2011	dlc	8204	Engenharia Electrónica e Telecomunicações	45996	136	0	5	5	5	4

Figura 17 – Excerto do ficheiro Excel com os dados agregados

3.7 Resultados

Após ter sido modelada, a plataforma foi implementada seguindo as especificações descritas.

Devido à confidencialidade dos dados obtidos, não foi possível disponibilizar a plataforma desenvolvida num servidor público.

Esta plataforma é composta, assim, por 3 páginas: A página de entrada, a página com os gráficos de barras e a página com os gráficos de linhas.

O gráfico de barras apresenta o número de alunos inscritos, aprovados e avaliados nas diversas unidades curriculares dos cursos do Departamento de Electrónica, Telecomunicações e Informática entre os anos letivos 2009/2010 e 2012/2013.

Esta página apresenta alguns critérios de seleção: curso, unidade curricular, ano letivo, ano curricular e valor (que corresponde à seleção entre número de alunos inscritos, aprovados ou avaliados). Não foram considerados os critérios de seleção semestre e ano letivo por não haver dados que o permitissem.

O utilizador pode obter diferentes visualizações consoante os critérios seleccionados. Com as diversas combinações consegue visualizar, por exemplo, o número de alunos inscritos nas unidades curriculares de um ano específico para um dado curso, assim como a evolução do número de alunos aprovados numa unidade curricular ao longo dos anos letivos ou até a diferença do número de alunos avaliados numa dada unidade curricular nos diferentes cursos em que essa unidade curricular é lecionada.

O gráfico de linhas apresenta o rácio entre o número de alunos inscritos e o número de alunos aprovados numa unidade curricular, o rácio entre o número de alunos inscritos e o número de alunos avaliados numa unidade curricular e o rácio entre o número de alunos avaliados e o número de alunos aprovados numa unidade curricular.

Nesta página são apresentados dois critérios de seleção: a unidade curricular e o curso, sendo que a unidade curricular tem de ser obrigatoriamente seleccionada.

O utilizador pode escolher entre visualizar os 3 rácios disponíveis para uma dada unidade curricular de um curso apenas ou de todos os cursos que lecionam a unidade curricular seleccionada de forma a ser possível comparar os rácios entre os diversos cursos.

A página de entrada contém uma descrição relativamente a cada uma das representações acompanhada de uma imagem ilustrativa, como se pode ver na Figura 18. As visualizações são apresentadas ao seleccionar o botão respetivo.

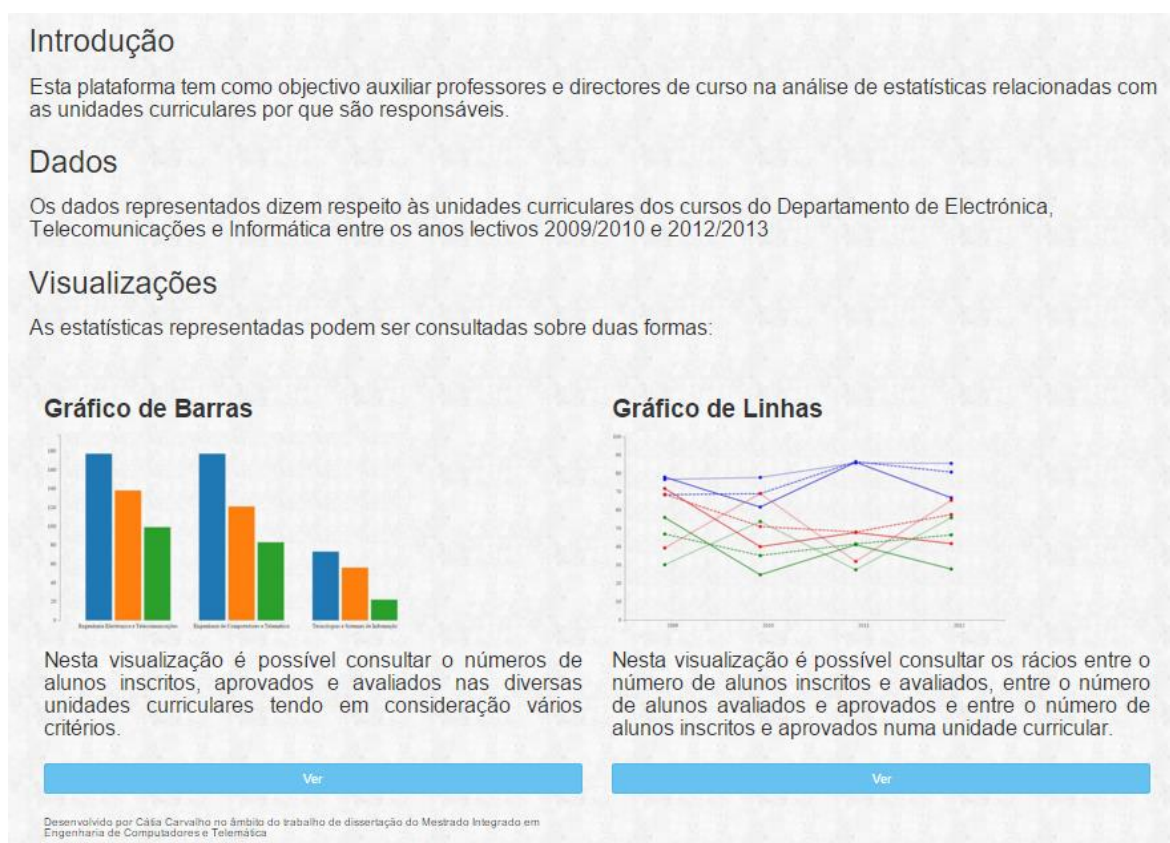


Figura 18 – Página de entrada da plataforma

Selecionando o botão para aceder ao gráfico de barras, na página de entrada, é apresentada a página ilustrada na Figura 19.



Figura 19 – Página do gráfico de barras sem filtros ativos

Nesta página o utilizador pode seleccionar os critérios que pretende. Ao seleccioná-los a visualização é apresentada, como mostra a Figura 20. Cada barra apresentada representa o número de alunos inscritos, avaliados ou aprovados numa unidade curricular de um curso num certo ano lectivo.

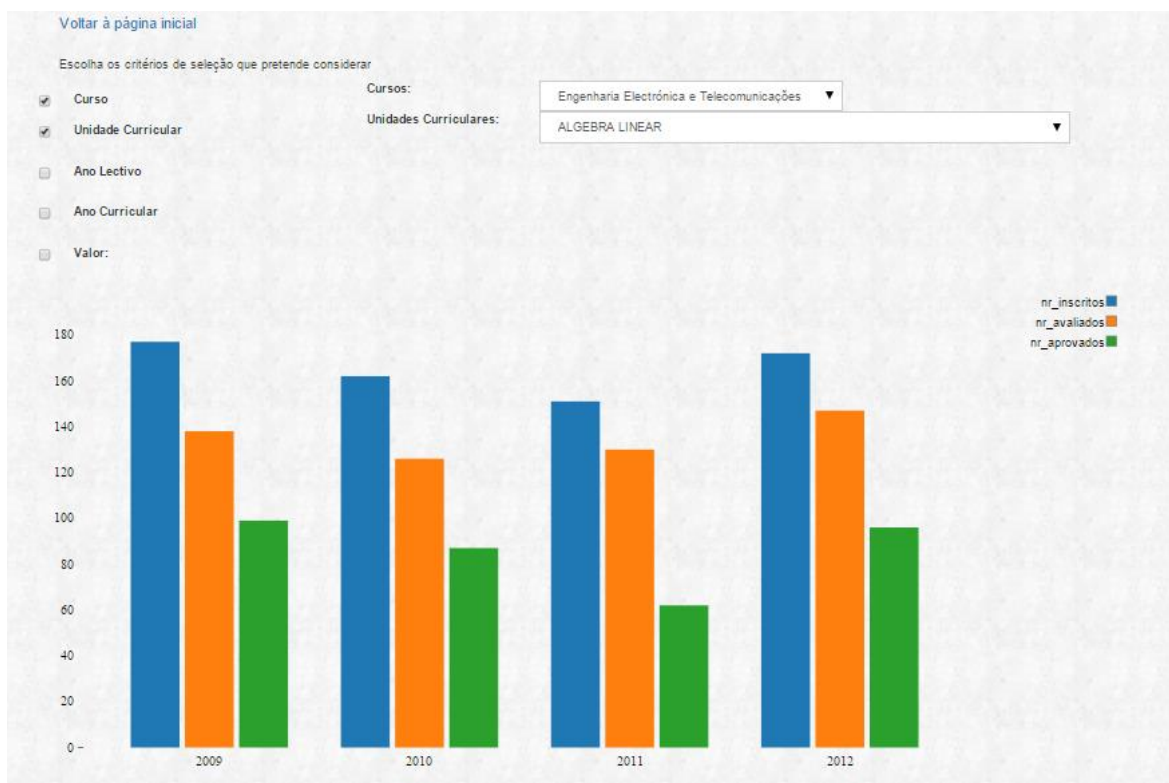


Figura 20 – Gráfico de barras

Ao passar o apontador por cima de qualquer barra é apresentada uma etiqueta com o número certo de alunos que essa barra representa, como se pode verificar na Figura 21.

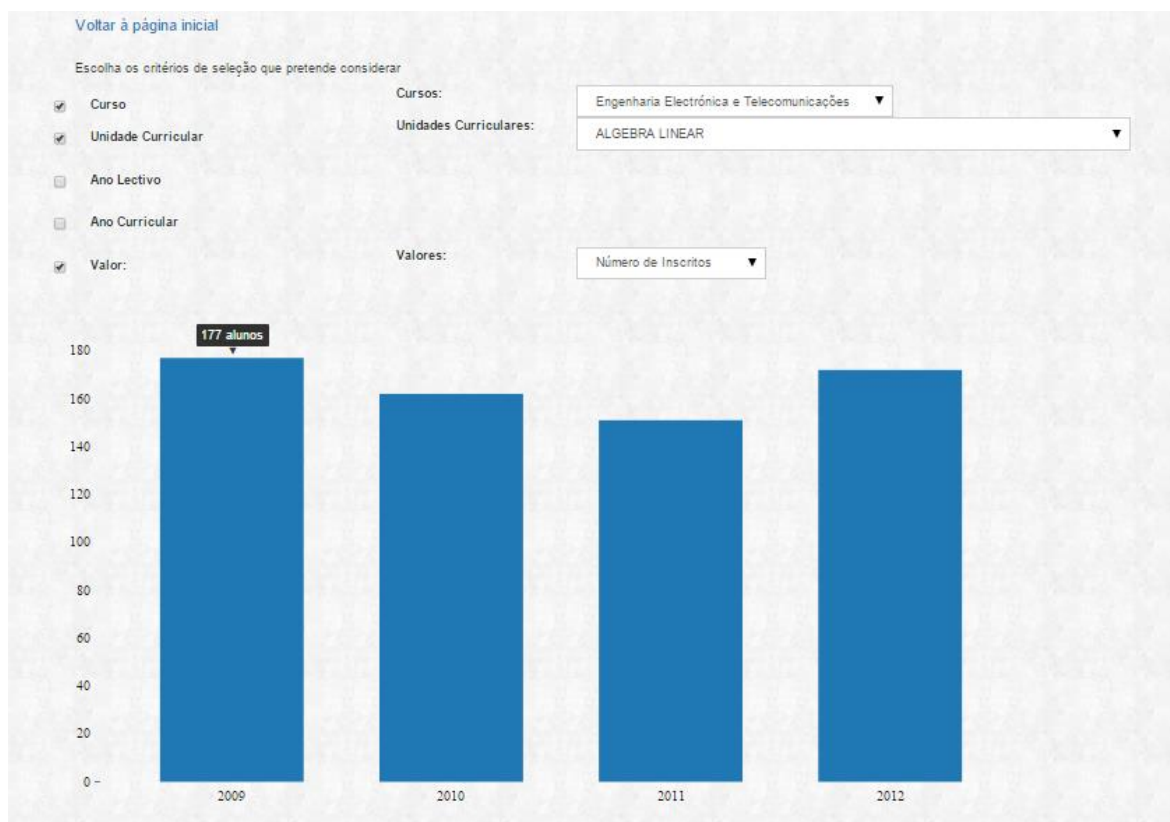


Figura 21 – Etiqueta apresentada no gráfico de barras

Caso seja selecionado um conjunto de critérios para o qual não existam dados ou a quantidade de dados seja muito grande é apresentada uma mensagem a indicar que não é possível apresentar uma visualização para esse conjunto de critérios, como se pode ver na Figura 22.



Figura 22 – Mensagem informativa na página do gráfico de barras

Quando as etiquetas dos eixos do gráfico são demasiado grandes ao ponto de se sobreporem, a visualização é gerada na horizontal e não na vertical, como mostra a Figura 23.

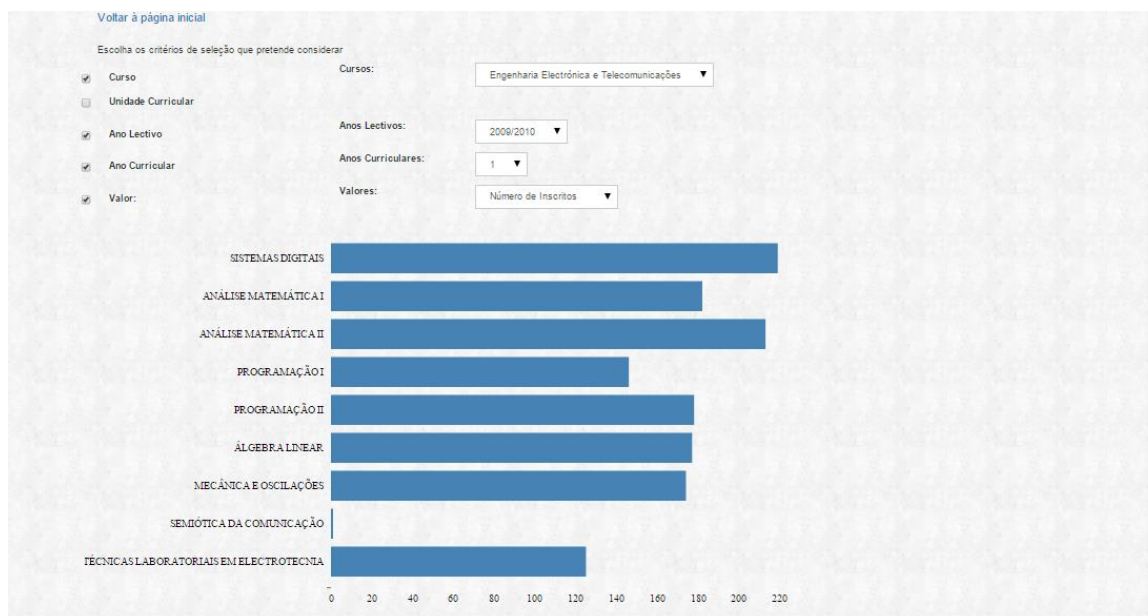


Figura 23 – Gráfico de barras na horizontal

Selecionando o botão para aceder ao gráfico de linhas, na página de entrada, é apresentada a página ilustrada na Figura 24. Cada linha representa a evolução do rácio entre o número de alunos aprovados e o número de alunos inscritos, do rácio entre o número de alunos avaliados e o número de alunos inscritos ou do rácio entre o número de alunos aprovados e o número de alunos avaliados para a unidade curricular seleccionada ao longo dos anos letivos em que essa unidade curricular foi lecionada.

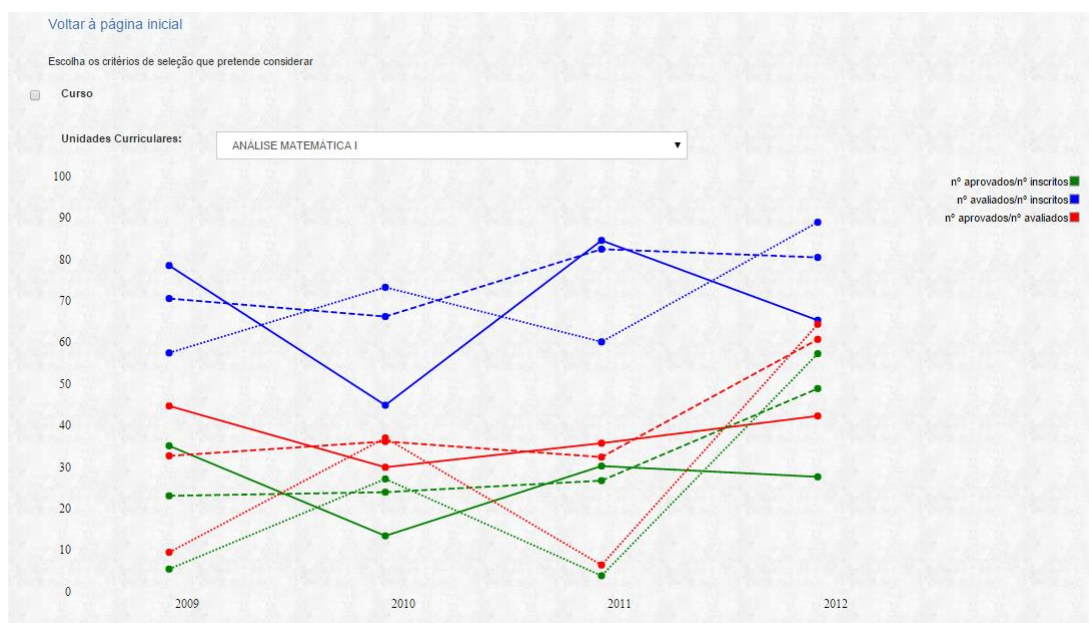


Figura 24 – Página do gráfico de linhas

Ao alterar a unidade curricular seleccionada o gráfico é atualizado automaticamente, como se pode verificar na Figura 25.



Figura 25 – Gráfico de linhas

Se o utilizador pretender visualizar os dados relativos a apenas um dos cursos, pode fazê-lo seleccionando o curso pretendido. O resultado obtido está representado na Figura 26.



Figura 26 – Gráfico de linhas para apenas um curso

À semelhança do gráfico de barras, o utilizador pode obter mais informações acerca da estatística passando o apontador por cima do ponto pretendido. A etiqueta apresentada esta representada na Figura 27.

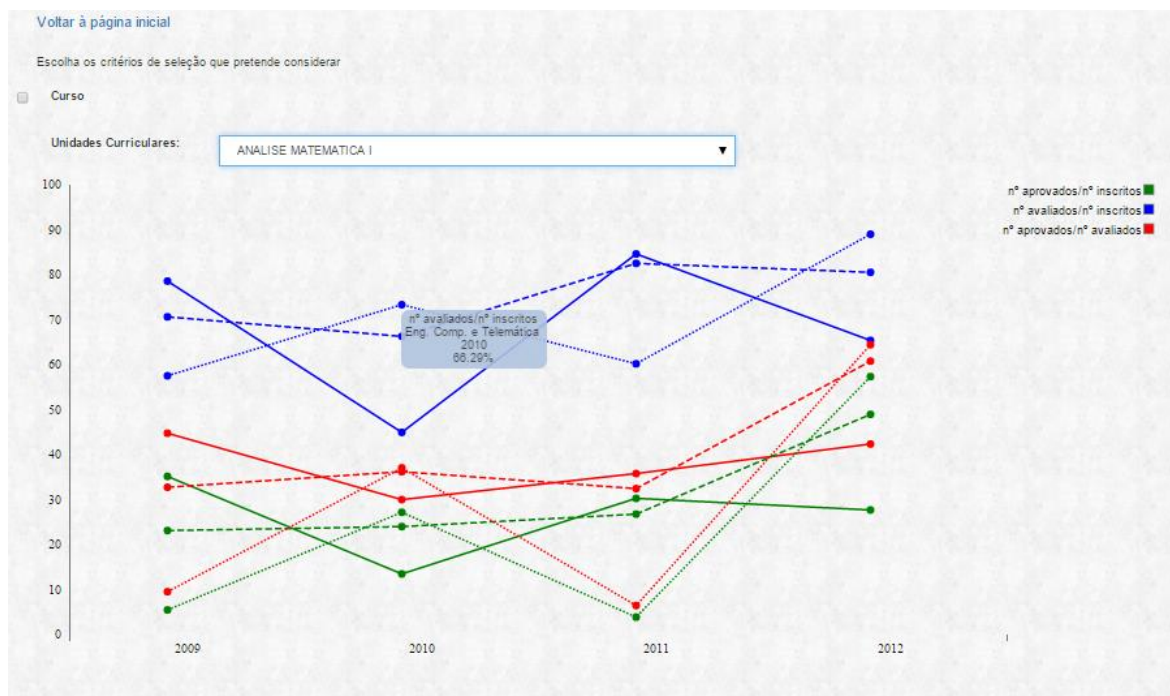


Figura 27 – Etiqueta apresentada no gráfico de linhas

Tanto na página do gráfico de barras como na página do gráfico de linhas o utilizador pode voltar à página inicial selecionando o link “Voltar à página inicial” que se encontra no canto superior esquerdo destas duas páginas.

4. Avaliação

Neste capítulo pretende-se descrever a avaliação efetuada à plataforma desenvolvida e apresentar os seus resultados. Para isso será feita uma avaliação heurística e de seguida testes com utilizadores.

4.1 Avaliação Heurística

A avaliação heurística consiste na verificação da existência de algumas características das visualizações, tendo em conta um conjunto de heurísticas.

Para a avaliação da plataforma desenvolvida neste trabalho de dissertação foram consideradas as heurísticas de Nielsen e Zuk. Não foram consideradas todas as heurísticas de Zuk uma vez que nem todas são relevantes para o trabalho desenvolvido.

4.1.1 Heurísticas de Nielsen

As heurísticas de Nielsen são 10 itens utilizados para avaliar a usabilidade de um website/plataforma web com o objetivo de evitar erros comuns que podem prejudicar a utilização do mesmo por parte de um utilizador [27].

As heurísticas são as seguintes:

1. Visibilidade do estado do sistema

O sistema deve informar continuamente o utilizador sobre o que está a acontecer com feedback apropriado a cada situação num intervalo de tempo razoável.

2. Falar a linguagem do utilizador

O sistema deve comunicar com o utilizador utilizando palavras, frases e conceitos familiares para o utilizador em vez de termos complexos. As informações devem ser organizadas conforme o modelo mental do utilizador.

3. Utilizador tem controlo e liberdade

Muitas vezes o utilizador seleciona uma função do sistema que na realidade não pretende e por isso necessita de ter uma opção de saída marcada de forma clara para que possa voltar ao estado anterior de forma simples. O sistema deve também ter a possibilidade de abortar uma tarefa ou desfazer uma operação a qualquer momento.

4. Consistência na utilização de normas

Uma operação deve estar sempre associada à mesma palavra, situação ou ação para facilitar o reconhecimento ao utilizador.

5. Evitar erros

Melhor do que gerar boas mensagens de erro é desenvolver um *design* cuidadoso que evita que o erro aconteça. Uma das formas de fazer isso é eliminar as condições que possam provocar esses erros ou até mesmo pedir que o utilizador confirme a sua escolha antes de efetuar a operação.

6. Reconhecer em vez de lembrar

O sistema deve mostrar os elementos de diálogo e permitir que o utilizador faça as suas escolhas, sem a necessidade de lembrar um comando específico.

7. Desenho simples e minimalista

As mensagens apresentadas ao utilizador não devem conter informações irrelevantes ou que não sejam necessárias.

8. Flexibilidade e eficiência

O sistema deve permitir o uso de atalhos para que utilizadores mais experientes e menos experientes tenham uma utilização mais eficiente.

9. Ajuda a reconhecer/diagnosticar/reparar erros

As mensagens de erro devem ser apresentadas numa linguagem clara e não em código. Devem indicar com precisão o problema e sugerir uma solução.

10. Ajuda e documentação

O ideal é que o sistema seja tão fácil de usar que não necessite de ajuda ou documentação, mas deve sempre fornecê-la.

Qualquer informação fornecida deve ser fácil de encontrar, deve ser focada na tarefa que o utilizador pretende efetuar, deve conter uma lista clara de passos a efetuar pelo utilizador e não deve ser muito extensa.

4.1.2 Heurísticas de Zuk et al.

As heurísticas de Zuk et al. são características utilizadas para a avaliação da usabilidade e apresentação de visualizações de informação. Estas heurísticas têm como objetivo evitar erros comuns e facilitar ao utilizador a leitura de uma visualização [28].

As heurísticas são as seguintes:

1. Certificar que a variável visual tem comprimento suficiente
2. Não esperar uma ordem de leitura através da cor
3. A percepção da cor varia com o tamanho do item colorido
4. O contraste local afeta a percepção da cor e de cinzento
5. Considerar pessoas daltónicas ou que não vêem cor
6. Os benefícios pré-atentivos aumentam com o campo de visão
7. Avaliação quantitativa requer variação de posição e tamanho
8. Conservar os dados para as dimensões do gráfico
9. Colocar a maior quantidade de dados num espaço pequeno
10. Remover o estranho (tinta)
11. Considerar as leis de Gestalt
12. Providenciar vários níveis de detalhe
13. Integrar texto onde for relevante

4.1.3 Avaliação

Considerando as heurísticas acima descritas foi feita a avaliação por dois utilizadores (1 do sexo masculino e outro do sexo feminino, ambos com bastante experiência na utilização de computadores) à plataforma desenvolvida atribuindo um grau de severidade a cada problema encontrado. Foi considerada a seguinte escala de gravidade:

- 0 – Mínimo
- 1 - Erro cosmético
- 2 - Problema menor
- 3 - Problema grave
- 4 - Erro catastrófico

Os problemas encontrados foram os seguintes:

- Não é dado qualquer tipo de feedback ao utilizador quando é escolhida uma combinação de critérios de seleção para a qual não existem resultados a mostrar.
 - Heurística de usabilidade violada - Visibilidade do estado do sistema
 - Severidade: 2
 - Sugestão: Mostrar uma mensagem informando o utilizador que não existem dados para a combinação de critérios de seleção escolhida.
- Não é apresentada a informação da página em que o utilizador se encontra.
 - Heurística de usabilidade violada - Visibilidade do estado do sistema
 - Severidade: 0
 - Sugestão: Acrescentar informação na página que informe o utilizador onde se encontra.
- Na legenda é utilizada uma abreviatura não muito conhecida para a palavra número (nr.).
 - Heurística de usabilidade violada – Falar a linguagem do utilizador
 - Severidade: 0
 - Sugestão: Substituir a abreviatura “nr.” pela abreviatura “nº” ou pela palavra “número”.

Conclui-se ainda que a plataforma cumpre as seguintes heurísticas:

- É possível alterar a combinação de critérios de seleção a qualquer momento e existe a possibilidade de voltar à página inicial sempre que o utilizador pretender.
 - Heurística de usabilidade cumprida – O utilizador tem controlo e liberdade
- A seleção de uma combinação de critérios de seleção leva sempre ao aparecimento de um gráfico.
 - Heurística de usabilidade cumprida – Consistência na utilização de normas
- Possui uma apresentação simples e clara.
 - Heurística de usabilidade cumprida – Desenho simples e minimalista
- Os formulários utilizados para apresentar os critérios de seleção são comuns.
 - Heurística de usabilidade cumprida – Reconhecer em vez de lembrar
- É possível avaliar as diferenças dos números representados apenas analisando a posição das barras/linhas nos gráficos apresentados.
 - Heurística de usabilidade cumprida - Avaliação quantitativa requer variação de posição e tamanho

- O tipo de gráfico apresentado ajusta-se ao conjunto de dados considerados.
 - Heurística de usabilidade cumprida – Conservar os dados para as dimensões do gráfico
- É possível obter mais ou menos detalhe aumentado ou diminuindo o número de critérios de seleção.
 - Heurística de usabilidade cumprida – Providenciar vários níveis de detalhe
- É apresentada informação detalhada ao passar com o cursor por cima do gráfico.
 - Heurística de usabilidade cumprida – Integrar texto onde for relevante

Após esta avaliação heurística foram corrigidos alguns dos problemas encontrados, como a falta de *feedback* ao utilizador quando não é possível apresentar a visualização para os critérios selecionados.

4.2 Testes de Usabilidade

Um dos métodos mais utilizados para avaliação em visualização de informação, para além da aplicação das heurísticas acima descritas, é a realização de testes de usabilidade.

Para a realização dos testes de usabilidade são desenvolvidas algumas tarefas que o utilizador deve tentar realizar após lhe ter sido apresentada e contextualizada a ferramenta. No final, o utilizador deve registar o grau de dificuldade sentido na realização da tarefa.

Para estes testes foram criados dois conjuntos de perguntas, um para cada tipo de visualização disponível. Todas as perguntas possuem um grau de dificuldade e objetivos equivalentes.

Cada utilizador é supervisionado por um observador, que regista o tempo necessário para a realização de cada tarefa assim como eventuais dificuldades encontradas. O observador regista ainda as observações feitas pelo utilizador.

Os questionários foram realizados por 10 utilizadores (8 do sexo masculino e 2 do sexo feminino com idades compreendidas entre os 20 e os 28 anos, todos com bastante experiência na utilização de computadores) que responderam a um total de 8 perguntas divididas pelas duas tarefas descritas acima. Cada tarefa era composta por 4 perguntas.

As perguntas e os formulários dos observadores são apresentados no Apêndice A.

Resultados

Após a realização dos testes observou-se que todos os utilizadores concluíram as duas tarefas sem errar nenhuma pergunta.

A Figura 28 mostra que o tempo de realização das tarefas varia entre os vários utilizadores, sendo o tempo médio de execução da primeira tarefa aproximadamente 4,5 minutos e da segunda tarefa 6 minutos. Também é possível observar que a Tarefa 2 demorou, em 80% dos casos, mais tempo a ser executada do que a Tarefa 1.

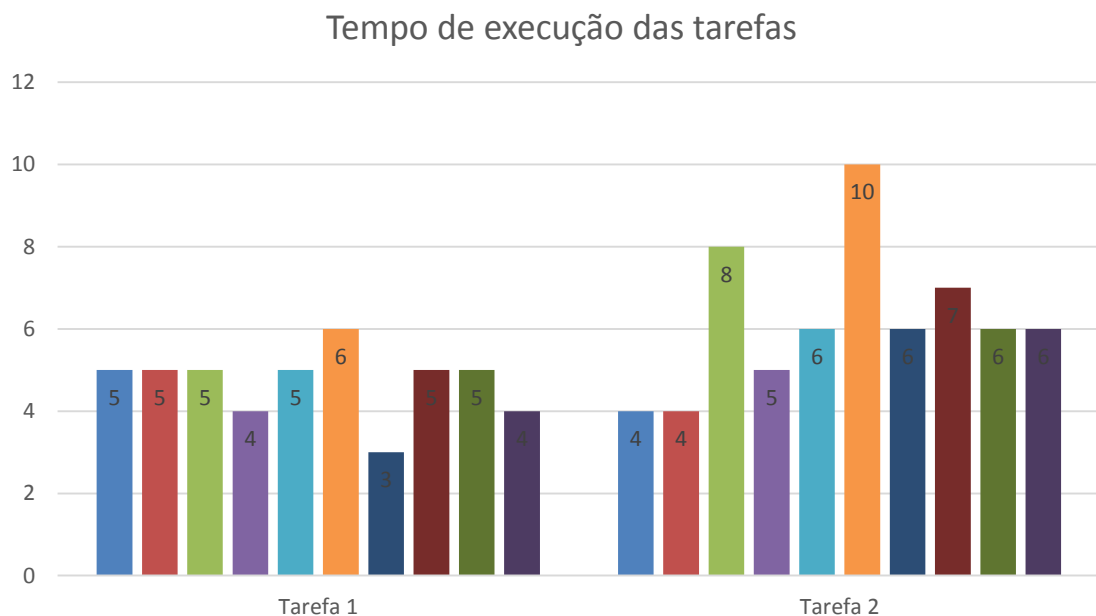


Figura 28 – Tempo de execução das tarefas por parte de cada utilizador

Foi pedido aos utilizadores que classificassem a dificuldade da execução de cada tarefa numa escala de 1 a 5, onde 1 significa muito difícil e 5 muito fácil.

A Figura 29 representa a dificuldade indicada pelos utilizadores na execução de cada tarefa.

Observando os dados apresentados é possível concluir que as duas tarefas apresentavam uma dificuldade semelhante, sendo a mediana da classificação de dificuldade de cada tarefa o valor 4.

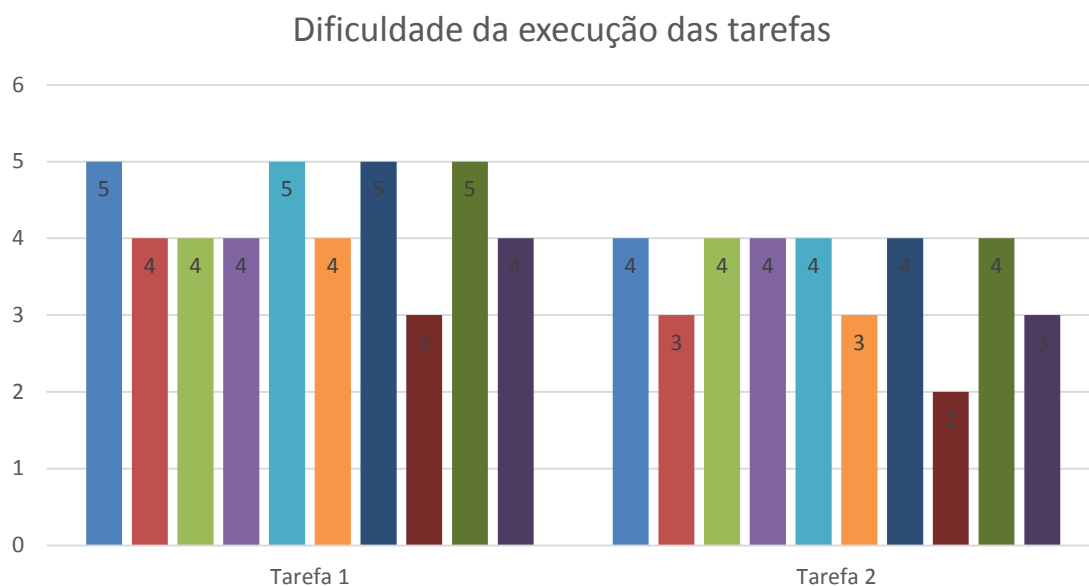


Figura 29 – Dificuldade indicada por cada utilizador na execução das tarefas

Durante as sessões de testes foram apontados pelos utilizadores comentários e críticas que, devido ao número reduzido de utilizadores, são mais importantes do que os resultados apresentados anteriormente.

Os comentários que mais se destacaram foram relativos ao gráfico de linhas (segunda tarefa).

Os utilizadores apontaram o facto de a legenda ser incompleta sendo necessário consultar a etiqueta apresentada para obter toda a informação, o problema das etiquetas serem apresentadas, por vezes, fora do sítio e ainda o facto de o ano letivo não ser apresentado corretamente nos eixos do gráfico sendo apenas apresentado o ano civil em que inicia o ano letivo e não a conotação usual onde é apresentado o ano civil em que inicia o ano letivo e o ano em que termina.

Os utilizadores apontaram ainda o facto do gráfico de linhas ser mais confuso do que o gráfico de barras por apresentar maior quantidade de dados.

5. Conclusões e trabalho futuro

Neste trabalho de dissertação foi desenvolvida uma plataforma com o objetivo de apresentar, através de representações visuais, os dados relativos ao número de alunos inscritos, avaliados e aprovados nas unidades curriculares dos cursos do Departamento de Electrónica, Telecomunicações e Informática.

Este capítulo irá resumir o objetivo do trabalho, impacto do trabalho desenvolvido e, finalmente, apresentar algumas sugestões para trabalho futuro.

5.1 *Resumo*

Este trabalho teve como objetivo principal desenvolver visualizações de informação capazes de representar taxas de abandono e de sucesso nas diversas unidades curriculares, utilizando dados existentes em ficheiros Excel e possibilitar o acesso a estes dados aos diretores de curso do Departamento de Electrónica, Telecomunicações e Informática.

Nesta dissertação foram apresentadas as várias fases do desenvolvimento das visualizações, desde a organização dos dados, o desenvolvimento das visualizações até à sua avaliação.

Para a execução destas fases da forma mais adequada foi necessário o acesso a literatura sobre a organização e armazenamento de grande quantidade de dados, assim como aceder a informação relativamente ao desenvolvimento das visualizações. Foi desenvolvido um repositório de dados e visualizações para apresentar os dados armazenados. Foi ainda necessário efetuar a avaliação da plataforma através de dois métodos: avaliação heurística e testes de usabilidade.

Assim, foram desenvolvidos dois tipos de visualização:

- Gráfico de barras – cada barra corresponde ao número de alunos inscritos, avaliados ou aprovados de uma dada unidade curricular.
- Gráfico de Linhas – cada linha representa a evolução da taxa de abandono ou de aprovação ao longo dos anos letivos em que a unidade curricular em questão foi lecionada.

Foi também apresentado um trabalho semelhante já existente. Nesse trabalho são também apresentados dados académicos em forma de representações visuais mas com o objetivo de encontrar relações entre as várias unidades curriculares em vez de analisar o desempenho dos alunos.

5.2 Trabalho futuro

Numa próxima fase, este trabalho pode ainda ser melhorado através da implementação das sugestões apontadas pelos utilizadores nos testes de usabilidade, podem ser corrigidos os restantes problemas encontrados na avaliação heurística da plataforma e através da implementação de novas funcionalidades, tais como, automatizar o carregamento dos dados na base de dados e implementação da visualização de dados relativos a motivos de abandono e entrada nos cursos.

No desenvolvimento das visualizações foram encontrados alguns problemas que poderão ser resolvidos também numa próxima fase, como por exemplo, a falta de uma segunda legenda e a posição das etiquetas no gráfico de linhas.

6. Bibliografia

- [1] T. Munzner, *Visualization Analysis and Design*. CRC Press, 2014.
- [2] E. Turban, J. E. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems*. Pearson/Prentice Hall, 2005.
- [3] C. W. H. F. Burstein, *Handbook on Decision Support Systems*. Springer-Verlag Berlin Heidelberg, 2008.
- [4] B. H. Boar, *Building, Using, and Managing the Data Warehouse-Understanding Data Warehousing Strategically*. Prentice Hall PTR, 1997.
- [5] V. Poe, P. Klauer, and S. Brobst, *Building a Data Warehouse for Decision Support*, 2nd Ed. Prentice Hall, 1998.
- [6] C. S. Jensen, T. B. Pedersen, and C. Thomsen, *Multidimensional databases and data warehousing*, vol. 2, no. 1. Morgan & Claypool, 2010.
- [7] "Data Warehouse Lifecycle Toolkit | Kimball Group." [Online]. Available: <http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-lifecycle-toolkit/>. [Accessed: 10-Dec-2014].
- [8] "Modelação de um Data Warehouse." [Online]. Available: <http://run.unl.pt/bitstream/10362/5366/1/TEGI0271.pdf>. [Accessed: 25-Sep-2015].
- [9] R. Kimball and M. Ross, *The Data Warehouse Toolkit : the complete guide to dimensional modeling*, 2nd ed. John Wiley and Sons, Inc., 2002.
- [10] "Kimball DW/BI Lifecycle Methodology | Kimball Group." [Online]. Available: <http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/>. [Accessed: 10-Dec-2014].
- [11] "Design Tip #115 Kimball Lifecycle in a Nutshell | Kimball Group." [Online]. Available: <http://www.kimballgroup.com/2009/08/design-tip-115-kimball-lifecycle-in-a-nutshell/>. [Accessed: 10-Dec-2014].
- [12] W. H. Inmon, *Building the Data Warehouse*, 3rd ed. John Wiley and Sons, Inc., 2002.
- [13] M. Breslin, *Data Warehousing Battle of the Giants*. 2004.
- [14] P. Corporation, "Pentaho Open Source Business Intelligence Platform: Technical White Paper," vol. 1, p. 20, 2006.
- [15] "Pentaho | Business analytics and business intelligence leaders." [Online]. Available: <http://www.pentaho.com/>. [Accessed: 06-Oct-2014].
- [16] "IBM Decision Management." IBM Corporation, 29-Sep-2015.
- [17] "IBM - Analytical Decision Management ." IBM Corporation, 06-Oct-2014.
- [18] "MICROSTRATEGY 9.4 . 1 VS . ORACLE 11G," 2014.
- [19] "Microstrategy 9.4.1 vOracle." [Online]. Available: <https://www.microstrategy.com/Strategy/media/downloads/white-papers/MicroStrategy-vs-Oracle-Deck.pdf>. [Accessed: 16-Oct-2015].
- [20] C. Chen, *Information Visualization*. Springer-Verlag London, 2006.
- [21] S. Liu, W. Cui, Y. Wu, and M. Liu, "A survey on information visualization: recent advances and challenges," Springer-Verlag Berlin Heidelberg, 2014, pp. 1373–1393.
- [22] "Charts | Google Developers." [Online]. Available: <https://google-developers.appspot.com/chart/>. [Accessed: 28-Sep-2015].

- [23] “Interactive JavaScript charts for your webpage | Highcharts.” [Online]. Available: <http://www.highcharts.com/>. [Accessed: 28-Sep-2015].
- [24] “Chart.js | Open source HTML5 Charts for your website.” [Online]. Available: <http://www.chartjs.org/>. [Accessed: 28-Sep-2015].
- [25] “D3.js - Data-Driven Documents.” [Online]. Available: <http://d3js.org/>. [Accessed: 28-Sep-2015].
- [26] S. Gama, V. Jordão, and D. Gonçalves, “EduVis: Visualização Interativa de Dados Educacionais,” Atas do 21º Encontro Português de Computação gráfica (EPCG 2014), pp. 25–32.
- [27] J. Nielsen, “10 Heuristics for User Interface Design: Article by Jakob Nielsen.” [Online]. Available: <http://www.nngroup.com/articles/ten-usability-heuristics/>. [Accessed: 02-Oct-2015].
- [28] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, and S. Carpendale, “Heuristics for Information Visualization Evaluation,” BELIV ’06 Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization, 2006, pp. 1–6.

Apêndice A Testes de Usabilidade: Documentos

Apêndice A.1 Guia de tarefas do Utilizador

Sucesso/Insucesso dos alunos

Sessão de Avaliação

Este documento apresenta uma **lista de tarefas** para serem executadas sobre as visualizações disponíveis. Este teste de usabilidade destina-se a avaliar o sistema e não o utilizador. Como tal, não se sinta pressionado pelo tempo nem pela necessidade em completar as tarefas com sucesso. Se sentir dificuldades, pode pedir ajuda ou desistir livremente de uma ou mais tarefas.

Nº Utilizador: _____

Idade: _____

Sexo: M ☐ F ☐

Escolaridade: _____

Indique a sua experiência na utilização de um computador.

nada experiente

muito experiente.

1 ☐

2 ☐

3 ☐

4 ☐

5 ☐

Tarefa 1	<p>Para o gráfico de barras disponível a partir da página inicial.</p> <p>Observando e interagindo com a representação do número de alunos inscritos, avaliados e aprovados nas várias unidades curriculares dos diversos cursos, responda às seguintes questões:</p> <p>A) Qual o número de alunos avaliados na unidade curricular “Computação Visual” no ano letivo 2013/2014?</p> <p>Resposta: _____</p> <p>B) Qual a unidade curricular do 1º ano do curso de Engenharia de Computadores e Telemática com maior número de alunos inscritos no ano letivo 2010/2011?</p> <p>Resposta: _____</p> <p>C) Qual a diferença entre o número de alunos inscritos e avaliados na unidade curricular “Bases de dados” no ano letivo 2012/2013?</p> <p>Resposta: _____</p> <p>D) Em que ano letivo houve o maior número de alunos de Engenharia de Electrónica e Telecomunicações inscritos na unidade curricular “Electrónica I”?</p> <p>Resposta: _____</p>								
	<table border="1"> <tr> <td>Facilidade da Tarefa 1</td> <td>Nada Fácil</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>Muito Fácil</td> </tr> </table>	Facilidade da Tarefa 1	Nada Fácil	1	2	3	4	5	Muito Fácil
	Facilidade da Tarefa 1	Nada Fácil	1	2	3	4	5	Muito Fácil	

Tarefa 2	<p>Para o gráfico de linhas disponível a partir da página inicial.</p> <p>Observando e interagindo com a representação dos rácios entre o número de alunos inscritos, aprovados, avaliados numa unidade curricular, responda às seguintes questões:</p> <p>A) Qual a percentagem de alunos de Engenharia de Computadores e Telemática inscritos na unidade curricular “Arquitetura de Computadores I” que obteve aprovação no ano letivo 2010/2011?</p> <p>Resposta: _____</p> <p>B) Qual a percentagem de alunos de Engenharia de Eletrónica e Telecomunicações inscritos na unidade curricular “Análise de Circuitos” que foi avaliada no ano letivo 2011/2012?</p> <p>Resposta: _____</p> <p>C) Qual o curso com a maior percentagem de alunos avaliados na unidade curricular “Análise Matemática I” que teve aprovação no ano letivo 2009/2010?</p> <p>Resposta: _____</p> <p>D) Qual a diferença entre a percentagem de alunos avaliados do curso Engenharia de Computadores e Telemática e a percentagem de alunos avaliados do curso Engenharia de Eletrónica e Telecomunicações que obteve aprovação na unidade curricular “Análise Matemática 1” no ano letivo 2011/2012?</p> <p>Resposta: _____</p>								
	<table border="1"> <tr> <td>Facilidade da Tarefa 2</td> <td>Nada Fácil</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>Muito Fácil</td> </tr> </table>	Facilidade da Tarefa 2	Nada Fácil	1	2	3	4	5	Muito Fácil
	Facilidade da Tarefa 2	Nada Fácil	1	2	3	4	5	Muito Fácil	

Apêndice A.2 Ficha do Observador

Tarefa	Completo(a) a tarefa?	Visualização	Sentiu-se perdido / cometeu erros?	Solicitou ajuda?
1	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desistiu	barras	<input type="checkbox"/> Sim: _____ <input type="checkbox"/> Não	<input type="checkbox"/> Sim: _____ <input type="checkbox"/> Não
	Observações:			
	Hora Início / Fim ____:____/____:____			

Tarefa	Completo(a) a tarefa?	Visualização	Sentiu-se perdido / cometeu erros?	Solicitou ajuda?
2	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desistiu	linhas	<input type="checkbox"/> Sim: _____ <input type="checkbox"/> Não	<input type="checkbox"/> Sim: _____ <input type="checkbox"/> Não
	Observações:			
	Hora Início / Fim ____:____/____:____			